

NOUVELLE SOUCHE DE CORONAVIRUS ASSOCIE AU SRAS ET SES APPLICATIONS

La présente invention est relative à une nouvelle souche de coronavirus associé au syndrome respiratoire aigu sévère (SRAS), issue d'un prélèvement répertorié sous le n° 031589 et prélevé à Hanoi (Vietnam), à des molécules d'acide nucléique issues de son génome, aux protéines et peptides codés par lesdites molécules d'acide nucléique ainsi qu'à leurs applications, notamment en tant que réactifs de diagnostic et/ou comme vaccin.

Le coronavirus est un virus à ARN monocaténaire, de polarité positive, d'approximativement 30 kilobases qui se réplique dans le cytoplasme des cellules hôtes ; l'extrémité 5' du génome a une structure en coiffe et l'extrémité 3' comporte une queue polyA. Ce virus est enveloppé et comprend, à sa surface, des structures péplomériques dénommées spicules.

Le génome comprend les cadres ouverts de lecture ou ORF suivants, de son extrémité 5' vers son extrémité 3' : ORF1a et ORF1b correspondant aux protéines du complexe de transcription-réplication, et ORF-S, ORF-E, ORF-M et ORF-N correspondant aux protéines structurales S, E, M et N. Il comprend également des ORFs correspondant à des protéines de fonction inconnue codées par : la région située entre l'ORF-S et l'ORF-E et chevauchant cette dernière, la région située entre l'ORF-M et l'ORF-N, et la région incluse dans l'ORF-N.

La protéine S est une glycoprotéine membranaire (200-220 kDa) qui se présente sous la forme de spicules ou "Spike" émergeant de la surface de l'enveloppe virale. Elle est responsable de l'attachement du virus aux récepteurs de la cellule hôte et de l'induction de la fusion de l'enveloppe virale avec la membrane cellulaire.

La petite protéine d'enveloppe (E) également dénommée sM (*small membrane*) qui est une protéine trans-membranaire non glycosylée d'environ 10 kDa, est la protéine présente en plus faible quantité dans le virion. Elle joue un rôle moteur dans le processus de bourgeonnement des coronavirus qui se produit au niveau du compartiment intermédiaire dans le réticulum endoplasmique et l'appareil de Golgi

La protéine M ou protéine de matrice (25-30 kDa) est une glycoprotéine membranaire plus abondante qui est intégrée dans la particule virale par une

interaction M/E, tandis que l'incorporation de S dans les particules est dirigée par une interaction S/M. Elle semble être importante pour la maturation virale des coronavirus et pour la détermination du site au niveau duquel les particules virales sont assemblées.

5 La protéine N ou protéine de nucléocapside (45-50 kDa) qui est la plus conservée parmi les protéines structurales des coronavirus, est nécessaire pour encapsider l'ARN génomique puis pour diriger son incorporation dans le virion. Cette protéine est vraisemblablement également impliquée dans la réplication de l'ARN.

Lorsqu'une cellule hôte est infectée, le cadre de lecture (ORF) situé
10 en 5' du génome viral est traduit en une polyprotéine qui est clivée par les protéases virales et libère alors plusieurs protéines non-structurales telles que l'ARN-polymérase ARN dépendante (Rep) et l'ATPase hélicase (Hel). Ces deux protéines sont impliquées dans la réplication du génome viral ainsi que dans la génération de
15 transcrits qui sont utilisés dans la synthèse des protéines virales. Les mécanismes par lesquels ces ARNms sub-génomiques sont produits, ne sont pas complètement compris ; cependant des faits récents indiquent que les séquences de régulation de la transcription à l'extrémité 5' de chaque gène représentent des signaux qui régulent la transcription discontinue des ARNms sub-génomiques.

Les protéines de la membrane virale (protéines S, E et M) sont insérées
20 rées dans le compartiment intermédiaire, alors que l'ARN répliqué (brin +) s'assemble avec la protéine N (nucléocapside). Ce complexe protéine-ARN s'associe ensuite avec la protéine M incluse dans les membranes du réticulum endoplasmique et les particules virales se forment lorsque le complexe de la nucléocapside bourgeonne dans le réticulum endoplasmique. Le virus migre ensuite à travers le complexe du Golgi et
25 éventuellement sort de la cellule, par exemple par exocytose. Le site de l'attachement du virus à la cellule hôte se trouve au niveau de la protéine S.

Les coronavirus sont responsables de 15 à 30 % des rhumes chez l'Homme et d'infections respiratoires ou digestives chez les animaux, notamment le chat (FIPV : *Feline infectious peritonitis virus*), la volaille (IBV : *Avian Infectious*
30 *bronchitis virus*), la souris (MHV : *Mouse Hepatitis virus*), le porc (TGEV : *Transmissible gastroenteritis virus*, PEDV : *Porcine Epidemic Diarrhea virus*,

PRCoV : *Porcine Respiratory Coronavirus*, HEV : *Hemagglutinating encephalomyelitis Virus*) et les bovins (BCoV : *Bovine coronavirus*).

En général, chaque coronavirus n'affecte qu'une seule espèce ; chez les individus immunocompétents, l'infection induit des anticorps éventuellement neutralisants et une immunité cellulaire, capables de détruire les cellules infectées.

Une épidémie de pneumonie atypique, dénommée syndrome respiratoire aigu sévère (SARS ou *Severe acute respiratory syndrome*, SRAS en français) s'est propagée dans différents pays (Vietnam, Hong-Kong, Singapour, Thaïlande et Canada) au cours du premier trimestre 2003, à partir d'un foyer initial apparu en Chine dans le dernier trimestre de 2002. La sévérité de cette maladie est telle que son taux de mortalité est d'environ 3 à 6 %. La détermination de l'agent causatif de cette maladie a été entreprise par de nombreux laboratoires, à travers le monde.

En mars 2003, un nouveau coronavirus (SARS-CoV, SARS virus ou virus SRAS, en français) a été isolé, en association avec des cas de syndrome respiratoire aigu sévère (T.G.KSIAZEK et al., *The New England Journal of Medicine*, 2003, 348, 1319-1330 ; C. DROSTEN et al., *The New England Journal of Medicine*, 2003, 348, 1967-1976 ; Peiris et al., *Lancet*, 2003, 361, 1319-).

Des séquences génomiques de ce nouveau coronavirus ont ainsi été obtenues, notamment celles de l'isolat Urbani (Genbank n° d'accès AY274119.3 et A. MARRA et al., *Science*, May 1, 2003, 300, 1399-1404) et de l'isolat de Toronto (Tor2, Genbank n° d'accès AY 278741 et A. ROTA et al., *Science*, 2003, 300, 1394-1399).

L'organisation du génome est comparable à celle des autres coronavirus connus permettant ainsi de confirmer l'appartenance du SARS-CoV à la famille des *Coronaviridae* ; les cadres ouverts de lecture ORF1a et 1b et les cadres ouverts de lecture correspondant aux protéines S, E, M, et N, ainsi qu'à des protéines codées par : la région située entre l'ORF-S et l'ORF-E (ORF3), la région située entre l'ORF-S et l'ORF-E et chevauchant l'ORF-E (ORF4), la région située entre l'ORF-M et l'ORF-N (ORF7 à ORF11) et la région correspondant à l'ORF-N (ORF13 et ORF14), ont notamment été identifiées.

Sept différences ont été mises en évidence entre les séquences des isolats Tor2 et Urbani ; 3 correspondent à des mutations silencieuses (c/t en position

16622 et a/g en position 19064 de l'ORF1b, t/c en position 24872 de l'ORF-S) et 4 modifient la séquence en acides aminés de respectivement : les protéines codées par l'ORF1a (c/t en position 7919 correspondant à la mutation A/V), la protéine S (g/t en position 23220 correspondant à la mutation A/S), la protéine codée par l'ORF3 (a/g en position 25298 correspondant à la mutation R/G) et de la protéine M (t/c en position 26857 correspondant à la mutation S/P).

En outre, l'analyse phylogénétique montre que le SARS-CoV est éloigné des autres coronavirus et qu'il est apparu, ni par mutation de coronavirus respiratoires humains, ni par recombinaison entre des coronavirus connus (pour une revue, voir Holmes, J.C.I., 2003, 111, 1605-1609).

La mise en évidence et la prise en compte de nouveaux variants sont importantes pour la mise au point de réactifs de détection et de diagnostic du SRAS suffisamment sensibles et spécifiques ainsi qu'à des compositions immunogènes aptes à protéger des populations contre des épidémies de SRAS.

Les Inventeurs ont maintenant mis en évidence une autre souche de coronavirus associé au SRAS, qui se distingue des isolats Tor2 et Urbani.

La présente invention a donc pour objet, une souche isolée ou purifiée de coronavirus humain associé au syndrome respiratoire aigu sévère, caractérisée en ce que son génome présente sous la forme d'ADN complémentaire un codon sérine en position 23220-23222 du gène de la protéine S ou un codon glycine en position 25298-25300 du gène de l'ORF3, et un codon alanine en position 7918-7920 de l'ORF1a ou un codon sérine en position 26857-26859 du gène de la protéine M, lesdites positions étant indiquées en référence à la séquence Genbank AY274119.3.

Selon un mode de réalisation avantageux de ladite souche, l'équivalent ADN de son génome présente une séquence correspondant à la séquence SEQ ID NO : 1 ; cette souche de coronavirus est issue du prélèvement de lavage bronchoalvéolaire d'un patient atteint de SRAS, répertorié sous le n° 031589 et effectué à l'hôpital français de Hanoi (Vietnam).

Conformément à l'invention, ladite séquence SEQ ID NO :1 est celle de l'acide désoxyribonucléique correspondant à la molécule d'acide ribonucléique du génome de la souche isolée de coronavirus telle que définie ci-dessus.

La séquence SEQ ID NO : 1 se distingue de la séquence Genbank AY274119.3 (isolat Tor2) en ce qu'elle possède les mutations suivantes :

- g/t en position 23220 ; le codon alanine (gct) en position 577 de la séquence en acides aminés de la protéine S de Tor2 est remplacé par un codon sérine (tct),

- a/g en position 25298 ; le codon arginine (aga) en position 11 de la séquence en acides aminés de la protéine codée par l'ORF3 de Tor 2 est remplacé par un codon glycine (gga).

En outre, la séquence SEQ ID NO : 1 se distingue de la séquence Genbank AY278741 (isolat Urbani) en ce qu'elle possède les mutations suivantes :

- t/c en position 7919 ; le codon valine (ggt) en position 2552 de la séquence en acides aminés de la protéine codée par l'ORF1a est remplacé par un codon alanine (gct),

- t/c en position 16622 : cette mutation ne modifie pas la séquence en acides aminés des protéines codées par l'ORF1b (mutation silencieuse),

- g/a en position 19064 : cette mutation ne modifie pas la séquence en acides aminés des protéines codées par l'ORF1b (mutation silencieuse),

- c/t en position 24872 : cette mutation ne modifie pas la séquence en acides aminés de la protéine S, et

- c/t en position 26857 : le codon proline (ccc) en position 154 de la séquence en acides aminés de la protéine M est remplacé par un codon sérine (tcc).

En l'absence de mention particulière, les positions des séquences nucléotidiques et peptidiques sont indiquées en référence à la séquence Genbank AY274119.3.

La présente invention a également pour objet un polynucléotide isolé ou purifié, caractérisé en ce que sa séquence est celle du génome de la souche isolée de coronavirus telle que définie ci-dessus.

Selon un mode de réalisation avantageux dudit polynucléotide il présente la séquence SEQ ID NO : 1.

La présente invention a également pour objet un polynucléotide isolé ou purifié, caractérisé en ce que sa séquence hybride dans des conditions de forte stringence avec la séquence du polynucléotide tel que défini ci-dessus.

Les termes « isolé ou purifié » signifient modifié « par la main de l'homme » à partir de l'état naturel ; autrement dit si un objet existe dans la nature, il est dit isolé ou purifié s'il a été modifié ou extrait de son environnement naturel ou les deux. Par exemple, un polynucléotide ou une protéine/un peptide naturellement présent dans un organisme vivant n'est ni isolé, ni purifié ; en revanche le même polynucléotide ou protéine /peptide séparé des molécules coexistantes dans son environnement naturel, obtenu par clonage, amplification et/ou synthèse chimique est isolé au sens de la présente invention. De plus, un polynucléotide ou une protéine/peptide qui est introduit dans un organisme par transformation, manipulation génétique ou par toute autre méthode, est « isolé » même s'il est présent dans ledit organisme. Le terme purifié tel qu'utilisé dans la présente invention, signifie que les protéines /peptides selon l'invention sont essentiellement libres d'association avec les autres protéines ou polypeptides, comme l'est par exemple le produit purifié de la culture de cellules hôtes recombinantes ou le produit purifié à partir d'une source non-recombinante.

Au sens de la présente invention, on entend par conditions d'hybridation de forte stringence, des conditions de température et de force ionique choisies de telle manière qu'elles permettent le maintien de l'hybridation spécifique et sélective entre polynucléotides complémentaires.

A titre d'illustration, des conditions de forte stringence aux fins de définir les polynucléotides ci-dessus, sont avantageusement les suivantes : l'hybridation ADN-ADN ou ADN-ARN est réalisée en deux étapes : (1) préhybridation à 42°C pendant 3 heures en tampon phosphate (20 mM pH 7,5) contenant 5 x SSC (1 x SSC correspond à une solution 0,15 M NaCl + 0,015 M citrate de sodium), 50 % de formamide, 7 % de sodium dodécyl sulfate (SDS), 10 x Denhardt's, 5 % de dextran sulfate et 1 % d'ADN de sperme de saumon ; (2) hybridation pendant 20 heures à 42°C suivie de 2 lavages de 20 minutes à 20°C en 2 x SSC + 2 % SDS, 1 lavage de 20 minutes à 20°C en 0,1 x SSC + 0,1 % SDS. Le dernier lavage est pratiqué en 0,1 x SSC + 0,1 % SDS pendant 30 minutes à 60°C.

La présente invention a également pour objet un fragment représentatif du polynucléotide tel que défini ci-dessus, caractérisé en ce qu'il est susceptible d'être obtenu, soit par l'utilisation d'enzymes de restriction dont les sites de reconnaissance et de coupure sont présents dans ledit polynucléotide tel que défini ci-

dessus, soit par amplification à l'aide d'amorces oligonucléotidiques spécifiques dudit polynucléotide tel que défini ci-dessus, soit par transcription *in vitro*, soit par synthèse chimique.

Selon un mode de réalisation avantageux dudit fragment, il est sélectionné dans le groupe constitué par : l'ADNc correspondant à au moins un cadre ouvert de lecture (ORF) choisi parmi : ORF1a, ORF1b, ORF-S, ORF-E, ORF-M, ORF-N, ORF3, ORF4, ORF7 à ORF11, ORF13 et ORF14, et l'ADNc correspondant aux extrémités 5' ou 3' non-codantes dudit polynucléotide.

Selon une disposition avantageuse de ce mode de réalisation, ledit fragment présente une séquence sélectionnée dans le groupe constitué par :

- les séquences SEQ ID NO : 2 et 4 représentant l'ADNc correspondant à l'ORF-S qui code pour la protéine S,
- les séquences SEQ ID NO : 13 et 15 représentant l'ADNc correspondant à l'ORF-E qui code pour la protéine E,
- les séquences séquence SEQ ID NO : 16 et 18 représentant l'ADNc correspondant à l'ORF-M qui code pour la protéine M,
- les séquences SEQ ID NO : 36 et 38 représentant l'ADNc correspondant à l'ORF-N qui code pour la protéine N,
- les séquences représentant les ADNc correspondant respectivement : aux ORF1a et ORF1b (ORF1ab, SEQ ID NO : 31), aux ORF3 et ORF4 (SEQ ID NO : 7, 8), aux ORF 7 à 11 (SEQ ID NO : 19, 20), à l'ORF13 (SEQ ID NO : 32) et à l'ORF14 (SEQ ID NO : 34), et
- les séquences représentant les ADNc correspondant respectivement aux extrémités 5'(SEQ ID NO : 39 et 72) et 3' non-codantes (SEQ ID NO : 40, 73) dudit polynucléotide.

La présente invention a également pour objet un fragment de l'ADNc codant pour la protéine S, tel que défini ci-dessus, caractérisé en ce qu'il présente une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 5 et 6 (fragments Sa et Sb).

La présente invention a également pour objet un fragment de l'ADNc correspondant aux ORF1a et ORF1b tel que défini ci-dessus, caractérisé en ce

qu'il présente une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 41 à 54 (fragments L0 à L12).

La présente invention a également pour objet un fragment du polynuéotide tel que défini ci dessus, caractérisé en ce qu'il présente au moins 15 bases
5 ou paires de bases consécutives de la séquence du génome de ladite souche incluant au moins une de celles situées en position 7979, 16622, 19064, 23220, 24872, 25298 et 26857. De préférence, il s'agit d'un fragment de 20 à 2500 bases ou paires de bases, de manière préférée de 20 à 400.

Selon un mode de réalisation avantageux dudit fragment, il inclut au
10 moins un couple de bases ou de paires de bases correspondant aux positions suivantes : 7919 et 23220, 7919 et 25298, 16622 et 23220, 19064 et 23220, 16622 et 25298, 19064 et 25298, 23220 et 24872, 23220 et 26857, 24872 et 25298, 25298 et 26857.

La présente invention a également pour objet des amorces d'au
15 moins 18 bases aptes à amplifier un fragment du génome d'un coronavirus associé au SRAS ou de l'équivalent ADN de celui-ci.

Selon un mode de réalisation desdites amorces, elles sont sélectionnées dans le groupe constitué par :

- la paire d'amorces n° 1 correspondant respectivement aux positions
20 28507 à 28522 (amorce sens, SEQ ID NO : 60) et 28774 à 28759 (amorce anti-sens, SEQ ID NO : 61) de la séquence du polynuéotide tel que défini ci-dessus,

- la paire d'amorces n° 2 correspondant respectivement aux positions 28375 à 28390 (amorce sens, SEQ ID NO : 62) et 28702 à 28687 (amorce anti-sens, SEQ ID NO : 63) de la séquence du polynuéotide tel que défini ci-dessus., et

25 - la paire d'amorces constituée des amorces SEQ ID Nos 55 et 56.

La présente invention a également pour objet une sonde apte à détecter la présence du génome d'un coronavirus associé au SRAS ou d'un fragment de celui-ci, caractérisée en ce qu'elle est sélectionnée dans le groupe constitué par : les fragments tels que définis ci-dessus et les fragments correspondant aux positions
30 suivantes de la séquence du polynuéotide tel que défini ci-dessus : 28561 à 28586, 28588 à 28608, 28541 à 28563 et 28565 à 28589 (SEQ ID NO : 64 à 67).

Les sondes et amorces selon l'Invention peuvent être marquées directement ou indirectement par un composé radioactif ou non radioactif par des méthodes bien connues de l'Homme du Métier, afin d'obtenir un signal détectable et/ou quantifiable. Parmi les isotopes radioactifs utilisés, on peut citer le ^{32}P , le ^{33}P , le ^{35}S , le ^3H ou l' ^{125}I . Les entités non radioactives sont sélectionnées parmi les ligands tels que la biotine, l'avidine, la streptavidine, la digoxygénine, les haptènes, les colorants, les agents luminescents tels que les agents radioluminescents, chémoluminescents, bioluminescents, fluorescents, phosphorescents.

L'invention englobe les sondes et les amorces marquées dérivées des séquences précédentes.

De telles sondes et amorces sont utiles pour le diagnostic de l'infection par un coronavirus associé au SRAS.

La présente invention a également pour objet une méthode de détection d'un coronavirus associé au SRAS, à partir d'un échantillon biologique, laquelle méthode est caractérisée en ce qu'elle comprend au moins :

(a) l'extraction d'acides nucléiques présents dans ledit échantillon biologique,

(b) l'amplification d'un fragment de l'ORF-N par RT-PCR à l'aide d'une paire d'amorces telle que définie ci-dessus, et

(c) la détection par tout moyen approprié des produits d'amplifications obtenus en (b).

Les produits d'amplifications (amplicons) en (b) sont de 268 pb pour la paire d'amorces n° 1 et de 328 pb pour la paire d'amorces n°2.

Selon un mode de mise en œuvre avantageux dudit procédé, l'étape (b) de détection est réalisée à l'aide d'au moins une sonde correspondant aux positions 28561 à 28586, 28588 à 28608, 28541 à 28563 et 28565 à 28589 de la séquence du polynucléotide tel que défini ci-dessus.

De préférence, le génome du coronavirus associé au SRAS est détecté et éventuellement quantifié par PCR en temps réel, à l'aide de la paire d'amorces n°2 et des sondes correspondant aux positions 28541 à 28563 et 28565 à

28589 marquées avec des composés différents, notamment des agents fluorescents différents.

La RT-PCR en temps réel qui met en œuvre cette paire d'amorces et cette sonde est très sensible puisqu'elle permet de détecter 10^2 copies d'ARN et jusqu'à 5 10 copies d'ARN, elle est en outre fiable et reproductible.

L'invention englobe les polydésoxyribonucléotides et les polyribonucléotides simple-brin, double-brin et tripe-brin correspondant à la séquence du génome de la souche isolée de coronavirus et de ses fragments tels que définis ci-dessus, ainsi qu'à leurs séquences complémentaires, sens ou anti-sens, notamment les 10 ARN et les ADNc correspondant à la séquence du génome et de ses fragments tels que définis ci-dessus.

La présente invention englobe également les fragments d'amplification obtenus à l'aide d'amorces spécifiques du génome de la souche purifiée ou isolée tel que défini ci-dessus, notamment à l'aide d'amorces et de paires 15 d'amorces telles que définies ci-dessus, les fragments de restriction constitués par ou comprenant la séquence des fragments tels que définis ci-dessus, les fragments obtenus par transcription *in vitro* à partir d'un vecteur contenant la séquence SEQ ID NO : 1 ou un fragment tel que défini ci-dessus, ainsi que des fragments obtenus par synthèse chimique. Des exemples de fragments de restriction sont déduits de la carte 20 de restriction de la séquence SEQ ID NO : 1 illustrée par la figure 13. Conformément à l'invention lesdits fragments sont, soit sous forme de fragments isolés, soit sous forme de mélanges de fragments. L'invention englobe également les fragments modifiés, par rapport aux précédents, par enlèvement, ou addition de nucléotides dans une proportion d'environ 15 %, par rapport à la longueur des fragments ci-dessus et/ou 25 modifiés au niveau de la nature des nucléotides, dès lors que les fragments nucléotidiques modifiés conservent une capacité d'hybridation avec les séquences d'ARN génomiques ou antigénomiques de l'isolat tel que défini ci-dessus.

Les molécules d'acide nucléique selon l'invention sont obtenues par les méthodes classiques, connues en elles-mêmes, en suivant les protocoles standards 30 tels que ceux décrits dans *Current Protocols in Molecular Biology* (Frederick M. AUSUBEL, 2000, Wiley and son Inc, Library of Congress, USA). Par exemple, elles

peuvent être obtenues par amplification d'une séquence nucléique par PCR ou RT-PCR ou bien par synthèse chimique totale ou partielle.

La présente invention a également pour objet une puce ou filtre à ADN ou à ARN, caractérisé en ce qu'il comprend au moins un polynucléotide ou l'un
5 de ses fragments tels que définis ci-dessus.

Les puces ou filtres à ADN ou à ARN selon l'invention sont préparés par les méthodes classiques, connues en elles-mêmes, comme par exemple greffage chimique ou électrochimique d'oligonucléotides sur support de verre ou de nylon.

10 La présente invention a également pour objet un vecteur de clonage et/ou d'expression recombinant, notamment un plasmide, un virus, un vecteur viral ou un phage comprenant un fragment d'acide nucléique tel que défini ci-dessus. De préférence, ledit vecteur recombinant est un vecteur d'expression dans lequel ledit fragment d'acide nucléique est placé sous le contrôle d'éléments régulateurs de la
15 transcription et de la traduction appropriés. En outre, ledit vecteur peut comprendre des séquences (étiquettes ou *tag*) fusionnées en phase avec l'extrémité 5' et/ou 3' dudit insert, utiles pour l'immobilisation, et/ou la détection et/ou la purification de la protéine exprimée à partir dudit vecteur.

Ces vecteurs sont construits et introduits dans des cellules hôtes par
20 les méthodes classiques d'ADN recombinant et de génie génétique, qui sont connues en elles-mêmes. De nombreux vecteurs dans lesquels on peut insérer une molécule d'acide nucléique d'intérêt afin de l'introduire et de la maintenir dans une cellule hôte, sont connus en eux-mêmes ; le choix d'un vecteur approprié dépend de l'utilisation envisagée pour ce vecteur (par exemple répllication de la séquence d'intérêt, expres-
25 sion de cette séquence, maintien de la séquence sous forme extrachromosomique ou bien intégration dans le matériel chromosomique de l'hôte), ainsi que de la nature de la cellule hôte.

Conformément à l'invention, ledit plasmide est notamment sélectionné parmi les plasmides suivants :

30 - le plasmide, dénommé SARS-S, compris dans la souche bactérienne déposée sous le n° I-3059, le 20 juin 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il

contient la séquence d'ADNc codant pour la protéine S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, laquelle séquence correspondant aux nucléotides des positions 21406 à 25348 (SEQ ID NO : 4), en référence à la séquence Genbank AY274119.3,

5 - le plasmide, dénommé SARS-S1, compris dans la souche bactérienne déposée sous le n° I-3020, le 12 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient un fragment 5' de la séquence d'ADNc codant pour la protéine S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, lequel fragment correspondant aux nucléotides des positions 21406 à 23454 (SEQ ID NO :5), en référence à la séquence Genbank AY274119.3 Tor2,

 - le plasmide, dénommé SARS-S2, compris dans la souche bactérienne déposée sous le n° I-3019, le 12 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il
15 contient un fragment 3' de la séquence d'ADNc codant pour la protéine S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, lequel fragment correspondant aux nucléotides des positions 23322 à 25348 (SEQ ID NO :6), en référence à la séquence Genbank n° d'accès AY274119.3,

 - le plasmide, dénommé SARS-SE, compris dans la souche bactérienne déposée sous le n° I-3126, le, 13 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15; il contient l'ADNc correspondant à la région située entre l'ORF-S et l'ORF-E et chevauchant l'ORF-E de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle région correspondant
25 aux nucléotides des positions 25110 à 26244 (SEQ ID NO :8), en référence à la séquence Genbank n° d'accès AY274119.3,

 - le plasmide, dénommé SARS-E, compris dans la souche bactérienne déposée sous le n° I-3046, le 28 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il
30 contient la séquence d'ADNc codant pour la protéine E de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle

séquence correspondant aux nucléotides des positions 26082 à 26413 (SEQ ID NO :15), en référence à la séquence Genbank n° d'accès AY274119.3,

- le plasmide, dénommé SARS-M ; compris dans la souche bactérienne déposée sous le n° I-3047, le 28 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc codant pour la protéine M de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus ; laquelle séquence correspondant aux nucléotides des positions 26330 à 27098 (SEQ ID NO :18), en référence à la séquence Genbank n° d'accès AY274119.3,

10 - le plasmide dénommé SARS-MN, compris dans la souche bactérienne déposée sous le n° I-3125, le 13 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc correspondant à la région située entre l'ORF-M et l'ORF-N de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589 et prélevée à Hanoi, telle que définie ci-dessus, laquelle séquence
15 correspondant aux nucléotides des positions 26977 à 28218 (SEQ ID NO :20), en référence à la séquence Genbank n° d'accès AY274119.3,

- le plasmide dénommé SARS-N, compris dans la souche bactérienne déposée sous le n° I-3048, le 5 juin 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il
20 contient l'ADNc codant pour la protéine N de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 28054 à 29430 (SEQ ID NO :38), en référence à la séquence Genbank n° d'accès AY274119.3 ; ainsi ce
25 plasmide comprend un insert de séquence SEQ ID NO :38 et est compris dans une souche bactérienne qui et qu'il a été déposée sous le n° I-3048, le 5 juin 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15,

- le plasmide dénommé SARS-5'NC, compris dans la souche bactérienne déposée sous le n° I- 3124, le 7 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient l'ADNc correspondant à l'extrémité 5'non codante du génome
30

de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 1 à 204 (SEQ ID NO :39), en référence à la séquence Genbank n° d'accès AY274119.3,

- le plasmide dénommé SARS-3'NC, compris dans la souche bactérienne déposée sous le n° I-3123 le 7 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15. ; il contient la séquence d'ADNc correspondant à l'extrémité 3'non codante du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant à celle située entre le nucléotide en position 28933 à 29727 (SEQ ID NO :40), en référence à la séquence Genbank n° d'accès AY274119.3, se termine par une série de nucléotides a.,

- le plasmide d'expression dénommé pIV2.3N, contenant un fragment d'ADNc codant pour une fusion C-terminale de la protéine N (SEQ ID NO : 37) avec une étiquette polyhistidine,

- le plasmide d'expression dénommé pIV2.3S_C, contenant un fragment d'ADNc codant pour une fusion C-terminale du fragment correspondant aux positions 475 à 1193 de la séquence en acides aminés de la protéine S (SEQ ID NO : 3) avec une étiquette polyhistidine,

- le plasmide d'expression pIV2.3S_L, contenant un fragment d'ADNc codant pour une fusion C-terminale du fragment correspondant aux positions 14 à 1193 de la séquence en acides aminés de la protéine S (SEQ ID NO : 3) avec une étiquette polyhistidine,

- le plasmide d'expression dénommé pIV2.4N, contenant un fragment d'ADNc codant pour une fusion N-terminale de la protéine N (SEQ ID NO : 3) avec une étiquette polyhistidine,

- le plasmide d'expression dénommé pIV2.4S_C ou pIV2.4S_I, contenant un insert codant pour une fusion N-terminale du fragment correspondant aux positions 475 à 1193 de la séquence en acides aminés de la protéine S (SEQ ID NO : 3) avec une étiquette polyhistidine, et

- le plasmide d'expression dénommé pIV2.4S_L contenant un fragment d'ADNc codant pour une fusion N-terminale du fragment correspondant aux

positions 14 à 1193 de la séquence en acides aminés de la protéine S (SEQ ID NO : 3) avec une étiquette polyhistidine.

Selon une disposition avantageuse du plasmide d'expression tel que défini ci-dessus, il est compris dans une souche bactérienne qui a été déposée sous le
5 n° I- 3117, le 23 octobre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15.

Selon une autre disposition avantageuse du plasmide d'expression tel que défini ci-dessus, il est compris dans une souche bactérienne qui a été déposée sous le n° I- 3118, le 23 octobre 2003, auprès de la Collection Nationale de Cultures
10 de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15.

Selon une autre disposition du plasmide d'expression tel que défini ci-dessus, il est compris dans une souche bactérienne qui a été déposée à la CNCM, 25 rue du Docteur Roux, 75724 Paris Cedex 15 sous les numéros suivants :

- a) souche n° I- 3118, déposée le 23 octobre 2003,
- 15 b) souche n° I- 3019, déposée le 12 mai 2003,
- c) souche n° I- 3020, déposée le 12 mai 2003,
- d) souche n° I-3059, déposée le 20 juin 2003,
- e) souche n° I-3323, déposée le 22 novembre 2004,
- f) souche n° I-3324, déposée le 22 novembre 2004,
- 20 g) souche n° I-3326, déposée le 1^{er} décembre 2004,
- h) souche n° I-3327, déposée le 1^{er} décembre 2004,
- i) souche n° I-3332, déposée le 1^{er} décembre 2004,
- j) souche n° I-3333, déposée le 1^{er} décembre 2004,
- k) souche n° I-3334, déposée le 1^{er} décembre 2004,
- 25 l) souche n° I-3335, déposée le 1^{er} décembre 2004,
- m) souche n° I-3336, déposée le 1^{er} décembre 2004,
- n) souche n° I-3337, déposée le 1^{er} décembre 2004,
- o) souche n° I-3338, déposée le 2 décembre 2004,
- p) souche n° I-3339, déposée le 2 décembre 2004,
- 30 q) souche n° I-3340, déposée le 2 décembre 2004,
- r) souche n° I-3341, déposée le 2 décembre 2004.

La présente invention a également pour objet un insérat d'acide nucléique d'origine virale, caractérisé en ce qu'il est contenu dans l'une quelconque des souches telles que définies ci-dessus en a)-r).

5 La présente invention a également pour objet un acide nucléique comportant un gène synthétique permettant une expression optimisée de la protéine S dans des cellules eucaryotes, caractérisé en ce qu'il possède la séquence SEQ ID NO: 140.

10 La présente invention a également pour objet un vecteur d'expression comportant un acide nucléique comportant un gène synthétique permettant une expression optimisée de la protéine S, lequel vecteur est contenu dans la souche bactérienne déposée à la CNCM, le 1^{er} décembre 2004, sous le n°I-3333.

Selon un mode de réalisation dudit vecteur d'expression, il s'agit d'un vecteur viral, sous forme de particule virale ou sous forme de génome recombinant.

15 Selon une disposition avantageuse de ce mode de réalisation, il s'agit d'une particule virale recombinante ou d'un génome viral recombinant susceptible d'être obtenu par transfection d'un plasmide selon les alinéas g), h) et k) à r) tels que définis ci-dessus, dans un système cellulaire approprié, c'est-à-dire, par exemple, des cellules transfectées avec un ou plusieurs autres plasmide(s), destinés à trancomplémenter certaines fonctions du virus délétées dans le vecteur et nécessaires à la formation des particules virales.

20 On entend ici par « famille de la protéine S » la protéine S complète, son ectodomaine et des fragments de cet ectodomaine qui sont de préférence produits dans un système eucaryote.

25 La présente invention a également pour objet un vecteur lentiviral codant pour un polypeptide de la famille de la protéine S, telle que définie ci-dessus.

La présente invention a également pour objet un virus rougeole recombinant codant pour un polypeptide de la famille de la protéine S, telle que définie ci-dessus.

30 La présente invention a également pour objet un virus vaccine recombinant codant pour un polypeptide de la famille de la protéine S, telle que définie ci-dessus.

La présente invention a également pour objet l'utilisation d'un vecteur selon les alinéas e) à r) tels que définis ci-dessus, ou d'un vecteur comportant exprimant un gène synthétique de la protéine S, tel que défini ci-dessus, pour la production, en système eucaryote, de la protéine S du coronavirus associé au SRAS, ou d'un fragment de cette protéine.

La présente invention a également pour objet une méthode de production de la protéine S en système eucaryote, comportant une étape de transfection de cellules eucaryotes en culture par un vecteur choisi parmi les vecteurs contenus dans les souches bactériennes mentionnées aux alinéas e) à r) ci-dessus ou un vecteur comportant un gène synthétique permettant une expression optimisée de la protéine S.

La présente invention a également pour objet une banque d'ADNc caractérisée en ce qu'elle comprend des fragments tels que définis ci-dessus, en particulier des fragments d'amplification ou des fragments de restriction, clonés dans un vecteur recombinant, notamment un vecteur d'expression (banque d'expression).

La présente invention a également pour objet des cellules, notamment des cellules procaryotes, modifiées par un vecteur recombinant tel que défini ci-dessus.

La présente invention a également pour objet une cellule eucaryote génétiquement modifiée exprimant une protéine ou un polypeptide tels que définis ci-après. Bien évidemment, les termes "cellule eucaryote génétiquement modifiée" ne désignent pas une cellule modifiée par un virus sauvage.

Selon un mode de réalisation avantageux de ladite cellule, elle est susceptible d'être obtenue par transfection par l'un quelconque des vecteurs mentionnés aux alinéas i) à l) ci-dessus.

Selon une disposition avantageuse de ce mode de réalisation, il s'agit de la cellule FRhK4-Ssol-30, déposée à la CNCM le 22 novembre 2004, sous le n°I-3325.

Les vecteurs recombinants tels que définis ci-dessus et les cellules transformées par lesdits vecteurs d'expression sont avantageusement utilisés pour la production des protéines et des peptides correspondants. Les banques d'expression dérivées desdits vecteurs, ainsi que les cellules transformées par lesdites banques

d'expression sont avantageusement utilisées pour identifier les épitopes immunogènes (épitopes B et T) des protéines du coronavirus associé au SRAS.

La présente invention a également pour objet les protéines et les peptides purifiées ou isolées, caractérisés en ce qu'ils sont codés par le polynucléotide
5 ou l'un de ses fragments tels que définis ci-dessus.

Selon un mode de réalisation avantageux de l'invention, ladite protéine est sélectionnée dans le groupe constitué par :

- la protéine S de séquence SEQ ID NO :3 ou son ectodomaine
- la protéine E de séquence SEQ ID NO :14
- 10 - la protéine M de séquence SEQ ID NO :17
- la protéine N de séquence SEQ ID NO : 37
- les protéines codées par les ORFs : ORF1a, ORF1b, ORF3, ORF4 et ORF7 à ORF11, ORF13 et ORF14 de séquence respectivement, SEQ ID NO :74, 75, 10, 12, 22, 24, 26, 28, 30, 33 et 35.

15 On utilisera ci-après indifféremment les termes « ectodomaine de la protéine S » et « forme soluble de la protéine S ».

Selon un mode de réalisation avantageux de l'invention, ledit peptide est constitué des acides aminés correspondant aux positions 1 à 1193 de la séquence en acides aminés de la protéine S

20 Selon un autre mode de réalisation avantageux de l'invention, ledit peptide est sélectionné dans le groupe constitué par :

- a) les peptides correspondant aux positions 14 à 1193 et 475 à 1193 de la séquence en acides aminés de la protéine S,
- b) les peptides correspondant aux positions 2 à 14 (SEQ ID NO : 69)
25 et 100 à 221 de la séquence en acides aminés de la protéine M ; ces peptides correspondent respectivement à l'ectodomaine et à l'endodomaine de la protéine M, et
- c) les peptides correspondant aux positions 1 à 12 (SEQ ID NO : 70) et 53 à 76 (SEQ ID NO : 71) de la séquence en acides aminés de la protéine E ; ces peptides correspondent respectivement à l'ectodomaine et à l'extrémité C-terminale de
30 la protéine E, et

d) les peptides de 5 à 50 acides aminés consécutifs, de préférence de 10 à 30 acides aminés, inclus ou chevauchant partiellement ou totalement la séquence des peptides tels que définis en a), b) ou c).

La présente invention a également pour objet un peptide caractérisé en ce qu'il présente une séquence de 7 à 50 acides aminés incluant un résidu d'acide aminé sélectionné dans le groupe constitué par :

- l'alanine située en position 2552 de la séquence en acides aminés de la protéine codée par l'ORF1a.

- la sérine située en position 577 de la séquence en acides aminés de la protéine S de la souche de SARS-CoV telle que définie ci-dessus,

- la glycine en position 11 de la séquence en acides aminés de la protéine codée par l'ORF3 de la souche de SARS-CoV telle que définie ci-dessus,

- la sérine en position 154 de la séquence en acides aminés de la protéine M de la souche de SARS-CoV telle que définie ci-dessus.

La présente invention a également pour objet un anticorps ou un fragment d'anticorps polyclonal ou monoclonal, susceptible d'être obtenu par immunisation d'un animal avec un vecteur recombinant tel que défini ci-dessus, une banque d'ADNc telle que définie ci-dessus ou bien une protéine ou un peptide tels que définis ci-dessus, caractérisé en ce qu'il se lie avec l'une au moins des protéines codées par le SARS-CoV telles que définies ci-dessus.

L'invention englobe les anticorps polyclonaux, les anticorps monoclonaux, les anticorps chimériques tels que les anticorps humanisés, ainsi que leurs fragments (Fab, Fv, scFv).

La présente invention a également pour objet un hybridome produisant un anticorps monoclonal contre la protéine N, caractérisé en ce qu'il est choisi parmi les hybridomes suivants :

- l'hybridome produisant l'anticorps monoclonal 87, déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-3328,

- l'hybridome produisant l'anticorps monoclonal 86, déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-3329,

- l'hybridome produisant l'anticorps monoclonal 57, déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-3330, et

- l'hybridome produisant l'anticorps monoclonal 156, déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-3331.

La présente invention a également pour objet un anticorps ou fragment d'anticorps polyclonal ou monoclonal dirigé contre la protéine N, caractérisé en ce qu'il est produit par un hybridome tel que défini ci-dessus.

Au sens de la présente invention, on entend par anticorps chimérique, relativement à un anticorps d'une espèce animale particulière ou d'une classe particulière d'anticorps, un anticorps comprenant tout ou partie d'une chaîne lourde et/ou d'une chaîne légère d'un anticorps d'une autre espèce animale ou d'une autre classe d'anticorps.

Au sens de la présente invention, on entend par anticorps humanisé une immunoglobuline humaine dans laquelle les résidus des CDRs (*Complementary-Determining Regions*) qui forment le site de liaison à l'antigène sont remplacés par ceux d'un anticorps monoclonal non-humain possédant la spécificité, l'affinité ou l'activité recherchées. Par comparaison avec les anticorps non-humains, les anticorps humanisés sont moins immunogènes et possèdent une demi-vie prolongée chez l'Homme car ils ne possèdent qu'une faible proportion de séquences non-humaines étant donné que la quasi-totalité des résidus des régions FR (Framework) et de la région constante (Fc) de ces anticorps sont ceux d'une séquence consensus d'immunoglobulines humaines.

La présente invention a également pour objet une puce ou filtre à protéine, caractérisé en ce qu'il comprend une protéine, un peptide ou bien un anticorps tels que définis ci-dessus.

Les puces à protéine selon l'invention sont préparées par les méthodes classiques, connues en elles-mêmes. Parmi les supports appropriés sur lesquels peuvent être immobilisés des protéines, on peut citer ceux en matière plastique ou en verre, notamment sous la forme de microplaques.

La présente invention a également pour objet des réactifs dérivés de la souche isolée de coronavirus associé au SRAS, issue du prélèvement répertorié sous le n° 031589, utiles pour l'étude et le diagnostic de l'infection provoquée par un coronavirus associé au SRAS, lesquels réactifs sont sélectionnés dans le groupe constitué par :

(a) une paire d'amorces, une sonde ou une puce à ADN telles que définies ci-dessus,

(b) un vecteur recombinant ou une cellule modifiée tels que définis ci-dessus,

5 (c) une souche isolée de coronavirus ou un polynucléotide tels que définis ci-dessus,

(d) une protéine ou un peptide tel que défini ci-dessus,

(e) un anticorps ou fragment d'anticorps tels que définis ci-dessus, et

(f) une puce à protéine telle que définie ci-dessus.

10 Ces différents réactifs sont préparés et utilisés selon les techniques classiques de biologie moléculaire et d'immunologie, en suivant les protocoles standards tels que ceux décrits dans *Current Protocols in Molecular Biology* (Frederick M. AUSUBEL, 2000, Wiley and Son Inc., Library of Congress, USA), dans *Current Protocols in Immunology* (John E. Cologan, 2000, Wiley and Son Inc.
15 Library of Congress, USA) et dans *Antibodies : A Laboratory Manual* (E. Howell and D Lane, Cold Spring Harbor Laboratory, 1988)

Les fragments d'acide nucléique selon l'invention sont préparés et utilisés selon les techniques classiques telles que définies ci-dessus. Les peptides et les protéines selon l'invention sont préparés par les techniques d'ADN recombinant,
20 connues de l'Homme du métier, notamment à l'aide des vecteurs recombinants tels que définis ci-dessus. Alternativement, les peptides selon l'invention peuvent être préparés par les techniques classiques de synthèse en phase solide ou liquide, connues de l'Homme du métier.

Les anticorps polyclonaux sont préparés par immunisation d'un
25 animal approprié avec une protéine ou un peptide tels que définis ci-dessus, éventuellement couplé à la KLH ou à l'albumine et/ou associé à un adjuvant approprié tel que l'adjuvant de Freund (complet ou incomplet) ou l'hydroxyde d'alumine ; après obtention d'un titre en anticorps satisfaisant, les anticorps sont récoltés par prélèvement du sérum des animaux immunisés et enrichis en IgG par précipitation, selon les
30 techniques classiques, puis les IgG spécifiques des protéines du SARS-CoV sont éventuellement purifiées par chromatographie d'affinité sur une colonne appropriée

sur laquelle sont fixés ledit peptide ou ladite protéine, tels que définis ci-dessus, de façon à obtenir une préparation d'IgG monospécifiques.

Les anticorps monoclonaux sont produits à partir d'hybridomes obtenus par fusion de lymphocytes B d'un animal immunisé par une protéine ou un peptide tels que définis ci-dessus avec des myélomes, selon la technique de Köhler et Milstein (Nature, 1975, 256, 495-497) ; les hybridomes sont cultivés *in vitro*, notamment dans des fermenteurs ou produits *in vivo*, sous forme d'ascite ; alternativement lesdits anticorps monoclonaux sont produits par génie génétique comme décrit dans le brevet américain US 4,816,567.

10 Les anticorps humanisés sont produits par des méthodes générales comme celles décrites dans la Demande Internationale WO 98/45332.

Les fragments d'anticorps sont produits à partir des régions V_H et V_L clonées, à partir des ARNm d'hybridomes ou de lymphocytes spléniques d'une souris immunisée ; par exemple, les fragments Fv, scFv ou Fab sont exprimés à la surface de phages filamenteux selon la technique de Winter et Milstein (Nature, 1991, 349, 293-299) ; après plusieurs étapes de sélection, les fragments d'anticorps spécifiques de l'antigène sont isolés et exprimés dans un système d'expression approprié, par les techniques classiques de clonage et d'expression d'ADN recombinant.

20 Les anticorps ou leur fragments tels que définis ci-dessus, sont purifiés par les techniques classiques connues de l'Homme du métier, telles que la chromatographie d'affinité.

La présente invention a en outre pour objet l'utilisation d'un produit sélectionné dans le groupe constitué par : une paire d'amorces, une sonde, une puce à ADN, un vecteur recombinant, une cellule modifiée, une souche isolée de coronavirus, un polynucléotide, une protéine ou un peptide, un anticorps ou un fragment d'anticorps et une puce à protéine tels que définis ci-dessus, pour la préparation d'un réactif de détection et éventuellement de génotypage/sérotypage, d'un coronavirus associé au SRAS.

30 Les protéines et les peptides selon l'invention, qui sont aptes à être reconnus et/ou à induire la production d'anticorps spécifiques du coronavirus associé au SRAS, sont utiles pour le diagnostic de l'infection par un tel coronavirus ; l'infection est détectée, par une technique appropriée- notamment EIA, ELISA, RIA,

immunofluorescence-, à partir d'un échantillon biologique prélevé chez un individu susceptible d'être infecté.

Selon une disposition avantageuse de ladite utilisation, lesdites protéines sont sélectionnées dans le groupe constitué par les protéines S, E, M et/ou N et les peptides tels que définis ci-dessus.

Les protéines S, E, M et/ou N et les peptides dérivés de ces protéines tels que définis ci-dessus, par exemple la protéine N, sont utilisées pour le diagnostic indirect d'une infection à coronavirus associé au SRAS (diagnostic sérologique ; détection d'anticorps spécifiques du SARS-CoV), notamment par une méthode immunoenzymatique (ELISA).

Les anticorps et les fragments d'anticorps selon l'invention, notamment ceux dirigés contre les protéines S, E, M et/ou N et les peptides dérivés tels que définis ci-dessus, sont utiles pour le diagnostic direct d'une infection à coronavirus associé au SRAS ; la détection de protéine(s) du SARS-CoV est réalisée par une technique appropriée, notamment EIA, ELISA, RIA, immunofluorescence à partir d'un échantillon biologique prélevé chez un individu susceptible d'être infecté.

La présente invention a également pour objet une méthode de détection d'un coronavirus associé au SRAS, à partir d'un échantillon biologique, laquelle méthode est caractérisée en ce qu'elle comprend au moins :

(a) la mise en contact dudit échantillon biologique avec au moins un anticorps ou un fragment d'anticorps, une protéine, un peptide ou bien une puce ou un filtre à protéine ou à peptide tels que définis ci-dessus, et

(b) la révélation par tout moyen approprié des complexes antigène-anticorps formés en (a), par exemple par EIA, ELISA, RIA, ou par immunofluorescence.

Selon un mode de mise en œuvre avantageux dudit procédé l'étape (a) comprend :

(a₁) la mise en contact dudit échantillon biologique avec au moins un premier anticorps ou un fragment d'anticorps qui est fixé sur un support approprié, notamment une microplaque,

(a₂) le lavage de la phase solide, et

(a₃) l'addition d'au moins un second anticorps ou un fragment d'anticorps, différent du premier, ledit anticorps ou fragment d'anticorps étant éventuellement marqué de façon appropriée.

Ce procédé qui permet de capturer les particules virales présentes
5 dans l'échantillon biologique est également dénommé procédé d'immunocapture.

Par exemple :

- l'étape (a₁) est réalisée avec au moins un premier anticorps monoclonal ou polyclonal ou un fragment de ceux-ci, dirigé contre la protéine S, M, et/ou E, et/ou un peptide correspondant à l'ectodomaine de l'une de ces protéines (peptides
10 M2-14 ou E1-12)

- l'étape (a₃) est réalisée avec au moins un anticorps ou un fragment d'anticorps dirigé contre un autre épitope de la même protéine ou de préférence contre une autre protéine, de manière préférée contre une protéine interne telle que la nucléoprotéine N ou l'endodomaine de la protéine E ou M, de manière encore plus préférée
15 il s'agit d'anticorps ou de fragments d'anticorps dirigés contre la protéine N qui est très abondante dans la particule virale ; lorsqu'un anticorps ou un fragment d'anticorps dirigé contre une protéine interne (N) ou contre l'endodomaine des protéines E ou M est utilisé, le dit anticorps est incubé en présence de détergent, comme le Tween 20 par exemple, à des concentrations de l'ordre de 0,1 %.

- l'étape (b) de révélation des complexes antigène-anticorps formés
20 est réalisée, soit directement à l'aide d'un second anticorps marqué par exemple avec de la biotine ou une enzyme appropriée telle que la peroxydase ou la phosphatase alcaline, soit indirectement à l'aide d'un sérum anti-immunoglobulines marqué comme ci-dessus. Les complexes ainsi formés sont révélés à l'aide d'un substrat
25 approprié.

Selon une mise en œuvre préférée de cet aspect de l'invention, l'échantillon biologique est mélangé à l'anticorps monoclonal de révélation préalablement à sa mise en contact avec les anticorps monoclonaux de capture. Le cas échéant, le mélange sérum-anticorps de révélation est incubé pendant au moins 10 minutes à
30 température ambiante avant d'être appliqué sur la plaque.

La présente invention a également pour objet un test d'immunocapture destiné à détecter une infection par le coronavirus associé au SRAS par

détection de la nucléoprotéine native (protéine N), en particulier caractérisé en ce que l'anticorps utilisé pour la capture de la nucléoprotéine virale native est un anticorps monoclonal spécifique de la région centrale et/ou d'un épitope conformationnel.

Selon un mode de réalisation dudit test, l'anticorps utilisé pour la
5 capture de la protéine N est l'anticorps monoclonal Acm87, produit par l'hybridome déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-3328.

Selon un autre mode de réalisation dudit test d'immunocapture, l'anticorps utilisé pour la capture de la protéine N est l'anticorps monoclonal Acm86, produit par l'hybridome déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-
10 3329.

Selon un autre mode de réalisation dudit test d'immunocapture, les anticorps monoclonaux Acm86 et Acm87 sont utilisés pour la capture de la protéine N.

Dans les tests d'immunocapture selon l'invention, on peut utiliser
15 pour la révélation de la protéine N, l'anticorps monoclonal Acm57, produit par l'hybridome déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-3330, ledit anticorps étant conjugué à une molécule ou une particule révélatrice.

Conformément audit test d'immunocapture, une combinaison des anticorps Acm57 et Acm87, conjugués à une molécule ou une particule révélatrice, est
20 utilisée pour la révélation de la protéine N.

Une molécule révélatrice peut être un atome radioactif, un colorant, une molécule fluorescente, un fluorophore, une enzyme ; une particule révélatrice peut être, par exemple : de l'or colloïdal, une particule magnétique ou une bille de latex.

La présente invention a également pour objet un réactif de détection
25 d'un coronavirus associé au SRAS, caractérisé en ce qu'il est sélectionné dans le groupe constitué par :

- (a) une paire d'amorces ou une sonde telles que définies ci-dessus,
- (b) un vecteur recombinant tel que défini ci-dessus ou une cellule modifiée telle que définie ci-dessus,
- 30 (c) une souche isolée de coronavirus telle que définie ci-dessus ou un polynucléotide tel que défini ci-dessus,
- (d) un anticorps ou un fragment d'anticorps tel que défini ci-dessus,

(e) une combinaison d'anticorps comprenant les anticorps monoclonaux Acm86 et/ou Acm87, et l'anticorps monoclonal Acm57, telle que définie ci-dessus,

(f) une puce ou un filtre tels que définis ci-dessus.

5 La présente invention a également pour objet une méthode de détection d'une infection par un coronavirus associé au SRAS, à partir d'un échantillon biologique, par ELISA IgG indirect utilisant la protéine N, laquelle méthode est caractérisée en ce que les plaques sont sensibilisées par une solution de protéine N à une concentration comprise entre 0,5 et 4 μ g/mL, de préférence 2 μ g/mL, dans un
10 tampon PBS 10mM pH 7,2, rouge de phénol à 0,25mL /L.

La présente invention a, en outre, pour objet une méthode de détection d'une infection par un coronavirus associé au SRAS, à partir d'un échantillon biologique, par ELISA double épitope, caractérisée en ce que le sérum à tester est mélangé à l'antigène de révélation, ledit mélange étant ensuite mis au contact de l'anti-
15 gène fixé sur un support solide.

Selon une variante des tests de détection des coronavirus associé au SRAS, ces tests combinent un ELISA utilisant la protéine N, et un autre ELISA utilisant la protéine S, tel que décrit plus bas.

La présente invention a aussi pour objet un complexe immunitaire formé
20 d'un anticorps ou d'un fragment d'anticorps polyclonal ou monoclonal tel que défini ci-dessus, et d'une protéine ou d'un peptide du coronavirus associé au SRAS.

La présente invention a en outre pour objet un kit de détection d'un coronavirus associé au SRAS, caractérisé en ce qu'il comprend au moins un réactif sélectionné dans le groupe constitué par : une paire d'amorces, une sonde, une puce à
25 ADN ou à ARN, un vecteur recombinant, une cellule modifiée, une souche isolée de coronavirus, un polynucléotide, une protéine ou un peptide, un anticorps, et une puce à protéine tels que définis ci-dessus.

La présente invention a en outre pour objet, une composition immunogène, caractérisée en ce qu'elle comprend au moins un produit sélectionné
30 dans le groupe constitué par :

a) une protéine ou un peptide tels que définis ci-dessus,

b) un polynucléotide de type ADN ou ARN ou l'un de ses fragments représentatifs tels que définis ci-dessus, de séquence choisie parmi :

(i) la séquence SEQ ID NO : 1 ou son équivalent ARN

(ii) la séquence hybridant dans des conditions de forte stringence
5 avec la séquence SEQ ID NO : 1,

(iii) la séquence complémentaire de la séquence SEQ ID NO : 1 ou de la séquence hybridant dans des conditions de forte stringence avec la séquence SEQ ID NO : 1,

(iv) la séquence nucléotidique d'un fragment représentatif du poly-
10 nucléotide tel que défini en (i), (ii) ou (iii),

(v) la séquence telle que définie en (i), (ii), (iii) ou (iv), modifiée, et

c) un vecteur d'expression recombinant comprenant un polynucléotide tel que défini en b), et

d) une banque d'ADNc telle que définie ci-dessus,

15 ladite composition immunogène étant capable d'induire une immunité humorale ou cellulaire protectrice spécifique du coronavirus associé au SRAS, notamment la production d'un anticorps dirigé contre un épitope spécifique du coronavirus associé au SRAS.

Les protéines et les peptides tels que définis ci-dessus, notamment
20 les protéines S, M, E et/ou N et les peptides dérivés, ainsi que les molécules d'acide nucléique (ADN ou ARN) codant lesdites protéines ou lesdits peptides, sont de bons candidats vaccin et peuvent être utilisées dans des compositions immunogènes pour la production d'un vaccin contre le coronavirus associé au SRAS.

Selon un mode de réalisation avantageux des compositions selon
25 l'invention, elles contiennent en outre, au moins un véhicule pharmaceutiquement acceptable et éventuellement des substances porteuses et/ou des adjuvants.

Les véhicules pharmaceutiquement acceptables, les substances porteuses et les adjuvants sont ceux classiquement utilisés.

Les adjuvants sont avantageusement choisis dans le groupe constitué
30 par des émulsions huileuses, de la saponine, des substances minérales, des extraits bactériens, de l'hydroxyde d'alumine et le squalène.

Les substances porteuses sont avantageusement sélectionnées dans le groupe constitué par des liposomes unilamellaires, des liposomes multilamellaires, des micelles de saponine ou des microsphères solides de nature saccharidique ou aurifère.

5 Les compositions selon l'invention, sont administrées par voie générale, notamment intramusculaire ou sous-cutanée ou bien par voie locale notamment nasale (aérosol).

La présente invention a également pour objet l'utilisation d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le
10 groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75 pour former un complexe immunitaire avec un anticorps dirigé spécifiquement contre un épitope du coronavirus associé au SRAS.

La présente invention a également pour objet un complexe immunitaire formé d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22,
15 24, 26, 28, 30, 33, 35, 37, 69, 70, 71, 74 et 75, et d'un anticorps dirigé spécifiquement contre un épitope du coronavirus associé au SRAS.

La présente invention a également pour objet l'utilisation d'une protéine ou d'un peptide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 3, 10, 12, 14, 17, 22, 24, 26, 28, 30,
20 33, 35, 37, 69, 70, 71, 74 et 75 pour induire la production d'un anticorps capable de reconnaître spécifiquement un épitope du coronavirus associé au SRAS.

La présente invention a également pour objet l'utilisation d'un polynuécléotide isolé ou purifié présentant une séquence sélectionnée dans le groupe constitué par les séquences SEQ ID NO : 1, 2, 4, 7, 8, 13, 15, 16, 18, 19, 20, 31, 36 et
25 38 pour induire la production d'un anticorps dirigé contre la protéine codée par ledit polynuécléotide et capable de reconnaître spécifiquement un épitope du coronavirus associé au SRAS

La présente invention a également pour objet des anticorps monoclonaux reconnaissant la protéine S native d'un coronavirus associé au SRAS.
30

La présente invention a également pour objet l'utilisation d'une protéine ou d'un polypeptide de la famille de la protéine S, telle que définie ci-dessus,

ou d'un anticorps reconnaissant la protéine S native, tel que défini ci-dessus, pour détecter une infection par un coronavirus associé au SRAS, à partir d'un échantillon biologique.

5 La présente invention a également pour objet une méthode de détection d'une infection par un coronavirus associé au SRAS, à partir d'un échantillon biologique, caractérisée en ce que la détection est effectuée par ELISA utilisant la protéine S recombinante, exprimée dans un système eucaryote.

Selon un mode de mise en œuvre avantageux de ladite méthode, il s'agit d'une méthode par ELISA double épitope, et le sérum à tester est mélangé à
10 l'antigène de révélation, ledit mélange étant ensuite mis au contact de l'antigène fixé sur un support solide.

La présente invention a aussi pour objet un complexe immunitaire formé d'un anticorps ou d'un fragment d'anticorps monoclonal reconnaissant la protéine S native, et d'une protéine ou d'un peptide du coronavirus associé au SRAS.

15 La présente invention a également pour objet un complexe immunitaire formé d'une protéine ou d'un polypeptide de la famille de la protéine S, telle que définie ci-dessus, et d'un anticorps dirigé spécifiquement contre un épitope du coronavirus associé au SRAS.

La présente invention a en outre pour objet un kit ou coffret de
20 détection d'un coronavirus associé au SRAS, caractérisé en ce qu'il comprend au moins un réactif sélectionné dans le groupe constitué par : une protéine ou polypeptide de la famille de la protéine S, telle que définie ci-dessus, un acide nucléique codant pour une protéine ou polypeptide de la famille de la protéine S, telle que définie ci-dessus, une cellule exprimant une protéine ou polypeptide de la famille de la protéine
25 S, telle que définie ci-dessus, ou un anticorps reconnaissant la protéine S native d'un coronavirus associé au SRAS.

La présente invention a pour objet une composition immunogène et/ou vaccinale, caractérisée en ce qu'elle comprend un polypeptide ou une protéine recombinante de la famille de la protéine S, telle que définie ci-dessus, obtenu dans un
30 système d'expression eucaryote.

La présente invention a également pour objet une composition immunogène et/ou vaccinale, caractérisée en ce qu'elle comprend un vecteur ou virus

recombinant, exprimant une protéine ou un polypeptide de la famille de la protéine S, telle que définie ci-dessus.

Outre les dispositions qui précèdent, l'invention comprend encore d'autres dispositions, qui ressortiront de la description qui va suivre, qui se réfère à des exemples de mise en œuvre du polynucléotide représentant le génome de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589, et des fragments d'ADNc dérivés objets de la présente invention, ainsi qu'au Tableau I présentant la liste des séquences :

Tableau I : Liste des séquences

numéro d'identification	Séquence	Position de l'ADNc en référence à Genbank AY274119.3	Numéro de dépôt à la CNCM du plasmide correspondant
SEQ ID NO : 1	génome de la souche issue du prélèvement 031589	-	-
SEQ ID NO : 2	ORF-S*	21406-25348	-
SEQ ID NO : 3	Protéine S	-	-
SEQ ID NO : 4	ORF-S**	21406-25348	I-3059
SEQ ID NO : 5	fragment Sa	21406-23454	I-3020
SEQ ID NO : 6	fragment Sb	23322-25348	I-3019
SEQ ID NO : 7	ORF-3+ORF-4*	25110-26244	-
SEQ ID NO : 8	ORF-3+ORF-4**	25110-26244	I-3126
SEQ ID NO : 9	ORF3	-	-
SEQ ID NO : 10	Protéine ORF-3	-	-
SEQ ID NO : 11	ORF4	-	-
SEQ ID NO : 12	Protéine ORF-4	-	-
SEQ ID NO : 13	ORF-E*	26082-26413	-
SEQ ID NO : 14	Protéine E	-	-
SEQ ID NO : 15	ORF-E**	26082-26413	I-3046
SEQ ID NO : 16	ORF-M*	26330-27098	-
SEQ ID NO : 17	Protéine M	-	-
SEQ ID NO : 18	ORF-M**	26330-27098	I-3047
SEQ ID NO : 19	ORF7 à 11*	26977-28218	-
SEQ ID NO : 20	ORF7 à 11**	26977-28218	I-3125
SEQ ID NO : 21	ORF7	-	-
SEQ ID NO : 22	Protéine ORF7	-	-
SEQ ID NO : 23	ORF8	-	-
SEQ ID NO : 24	Protéine ORF8	-	-
SEQ ID NO : 25	ORF9	-	-

SEQ ID NO : 26	Protéine ORF9	-	-
SEQ ID NO : 27	ORF10	-	-
SEQ ID NO : 28	Protéine ORF10	-	-
SEQ ID NO : 29	ORF11	-	-
SEQ ID NO : 30	Protéine ORF11	-	-
SEQ ID NO : 31	OrF1ab	265-21485	-
SEQ ID NO : 32	ORF13	28130-28426	-
SEQ ID NO : 33	Protéine ORF13	-	-
SEQ ID NO : 34	ORF14	-	-
SEQ ID NO : 35	Protéine ORF14	28583-28795	-
SEQ ID NO : 36	ORF-N*	28054-29430	
SEQ ID NO : 37	Protéine N	-	-
SEQ ID NO : 38	ORF-N**	28054-29430	I-3048
SEQ ID NO : 39	5'non-codante**	1-204	I-3124
SEQ ID NO : 40	3'non-codante**	28933-29727	I-3123
	ORF1ab		-
SEQ ID NO : 41	Fragment L0	30-500	
SEQ ID NO : 42	Fragment L1	211-2260	-
SEQ ID NO : 43	Fragment L2	2136-4187	-
SEQ ID NO : 44	Fragment L3	3892-5344	-
SEQ ID NO : 45	Fragment L4b	4932-6043	-
SEQ ID NO : 46	Fragment L4	5305-7318	-
SEQ ID NO : 47	Fragment L5	7275-9176	-
SEQ ID NO : 48	Fragment L6	9032-11086	-
SEQ ID NO : 49	Fragment L7	10298-12982	-
SEQ ID NO : 50	Fragment L8	12815-14854	-
SEQ ID NO : 51	Fragment L9	14745-16646	-
SEQ ID NO : 52	Fragment L10	16514-18590	-
SEQ ID NO : 53	Fragment L11	18500-20602	-
SEQ ID NO : 54	Fragment L12	20319-22224	-
SEQ ID NO : 55	Amorce N sens	-	-
SEQ ID NO : 56	Amorce N antisens	-	-
SEQ ID NO : 57	Amorce S _C sens	-	-
SEQ ID NO : 58	Amorce S _L sens	-	-
SEQ ID NO : 59	Amorce S _C e t S _L antisens	-	-
SEQ ID NO : 60	Amorce sens série 1	28507-28522	-
SEQ ID NO : 61	Amorce antisens série 1	28774-28759	
SEQ ID NO : 62	Amorce sens série 2	28375-28390	-
SEQ ID NO : 63	Amorce antisens série 2	28702-28687	-
SEQ ID NO : 64	Sonde 1/série 1	28561-28586	-

SEQ ID NO : 65	Sonde 2/série 1	28588-28608	-
SEQ ID NO : 66	Sonde 1/série 2	28541-28563	-
SEQ ID NO : 67	Sonde 2/série 2	28565-28589	-
SEQ ID NO : 68	Amorce ancre 14T		
SEQ ID NO : 69	Peptide M2-14	-	-
SEQ ID NO : 70	Peptide E1-12	-	-
SEQ ID NO : 71	Peptide E53-76	-	-
SEQ ID NO : 72	5'non-codante*	1-204	-
SEQ ID NO : 73	3'non-codante*	28933-29727	-
SEQ ID NO : 74	Protéine ORF1a	-	-
SEQ ID NO : 75	Protéine ORF1b	-	-
SEQ ID NO:76-139	Amorces		
SEQ ID NO:140	Pseudogène de S		
SEQ ID NO:141-148	amorces		
SEQ ID NO:149	Aa1-13 de S		
SEQ ID NO:150	polypeptide		
SEQ ID NO:151-158	amorces		

* produit d'amplification PCR (amplicon)

** insert cloné dans le plasmide déposé à la CNCM
ainsi qu'aux dessins annexés dans lesquels :

- la figure 1 illustre l'analyse par Western-blot de l'expression *in vitro* des protéines recombinantes N, S_C et S_L à partir des vecteurs d'expression pIVEX. Piste 1 : pIV2.3N. Piste 2 : pIV2.3S_C. Piste 3 : pIV2.3S_L. Piste 4 : pIV2.4N. Piste 5 : pIV2.4S₁ ou pIV2.4S_C. Piste 6 : pIV2.4S_L. L'expression de la protéine GFP exprimée à partir du même vecteur est utilisée comme contrôle.

- la figure 2 illustre l'analyse par électrophorèse en gel de polyacrylamide en conditions dénaturantes (SDS-PAGE) et coloration au bleu de Coomassie, de l'expression *in vivo* de la protéine N à partir des vecteurs d'expression pIVEX. La souche d'*E.coli* BL21(DE3)pDIA17 transformée par les vecteurs pIVEX recombinants est cultivée à 30°C dans du milieu LB, en présence ou en l'absence d'inducteur (IPTG 1mM). Piste 1 : pIV2.3N Piste 2 : pIV2.4N.

- la figure 3 illustre l'analyse par électrophorèse en gel de polyacrylamide en conditions dénaturantes (SDS-PAGE) et coloration au bleu de Coomassie, de l'expression *in vivo* des polypeptides S_L et S_C à partir des vecteurs

d'expression pIVEX. La souche d'*E.coli* BL21(DE3)pDIA17 transformée par les vecteurs pIVEX recombinants est cultivée à 30°C dans du milieu LB, en présence ou en l'absence d'inducteur (IPTG 1mM). Piste 1 : pIV2.3S_C Piste 2 : pIV2.3S_L. Piste 3 : pIV2.4S_I Piste 4 : pIV2.4S_L.

- 5 - la figure 4 illustre l'activité antigénique des protéines N, S_L et S_C recombinantes produites dans la souche *E. coli* BL21(DE3)pDIA17 transformée par les vecteurs pIVEX recombinants. A : électrophorèse (SDS-PAGE) des lysats bactériens. B et C : Western-blot avec les sérums, provenant d'un même patient infecté par le SARS-CoV, prélevés respectivement 8 jours (B : sérum M12) et 29 jours (C : sérum M13) après le début des symptômes du SRAS. Piste 1 : pIV2.3N. Piste 2 : pIV2.4N. Piste 3 : pIV2.3S_C. Piste 4 : pIV2.4 S_I. Piste 5 : pIV2.3S_L. Piste 6 : pIV2.4S_L

- 10 - la figure 5 illustre la purification sur colonne Ni-NTA agarose de la protéine N recombinante produite dans la souche *E. coli* BL21(DE3)pDIA17 à partir du vecteur pIV2.3N. Piste 1 : Extrait bactérien total. Piste 2 : Extrait soluble. Piste 3 : Extrait insoluble. Piste 4 : Extrait déposé sur la colonne Ni-NTA. Piste 5 : protéines non-retenues. Piste 6 : Fractions du pic 1. Piste 7 : Fractions du pic 2.

- 15 - la figure 6 illustre la purification de la protéine S_C recombinante à partir des corps d'inclusions produits dans la souche *E. coli* BL21(DE3)pDIA17 transformée par le pIV2.4S_I. A. Traitement au Triton X-100 (2%) : Piste 1 : Extrait bactérien total. Piste 2 : Extrait soluble. Piste 3 : Extrait insoluble. Piste 4 : Surnageant après traitement au Triton X-100 (2 %). Pistes 5 et 6 : Culot après traitement au Triton X-100 (2 %). B : Traitement à l'urée 4M, 5M, 6M et 7M des extraits solubles et insolubles.

- 20 - la figure 7 représente l'immunoempreinte réalisée à l'aide d'un lysat de cellules infectées par le SARS-CoV et d'un sérum de patient atteint de pneumopathie atypique.

- 25 - la figure 8 représente des immunoempreintes réalisées à l'aide d'un lysat de cellules infectées par le SARS-CoV et d'immunsérums de lapins spécifiques de la nucléoprotéine N (A) et de la protéine de spicule S (B). I.S. : sérum immun. p.i. : sérum pré-immun. L'immunsérum anti-N a été utilisé au 1/50000 et l'immunsérum anti-S au 1/10000.

- la figure 9 illustre la réactivité en ELISA des sérums polyclonaux monospécifiques de lapin dirigés contre la protéine N ou le fragment court de la protéine S (S_C), vis-à-vis des protéines recombinantes correspondantes utilisées pour l'immunisation. A : lapins P13097, P13081, et P13031 immunisés avec la protéine N recombinante purifiée. B : lapins P11135, P13042, et P14001 immunisés avec une
5 préparation de corps d'inclusions correspondants au fragment court de la protéine S (S_C). I.S. : sérum immun. p.i. : sérum pré-immun.

- la figure 10 illustre la réactivité en ELISA de la protéine N recombinante purifiée, vis-à-vis de sérum de patients atteints de pneumonie atypique causée par le SARS-CoV. Figure 10a : plaques ELISA préparés avec la protéine N à la
10 concentration de 4 µg/ml et 2 µg/ml. Figure 10b : plaque ELISA préparée avec la protéine N à la concentration de 1 µg/ml. Les sérums désignés A, B, D, E, F, G, H correspondent à ceux du Tableau IV.

- la figure 11 illustre l'amplification par RT-PCR de quantités décroissantes d'ARN synthétique du gène N du SARS-CoV (10^7 à 1 copie), à l'aide
15 des couples d'amorces n° 1 (N/+28507,N/-28774) (A) et n° 2 (N/+28375,N/-28702) (B). T : amplification réalisée en l'absence d'ARN. MW : marqueur d'ADN.

- la figure 12 illustre l'amplification par RT-PCR en temps réel d'ARN synthétique du gène N du SARS-CoV : des quantités décroissantes d'ARN
20 synthétique en répliquat (repli. ; pistes 16 à 29) ainsi que de l'ARN viral dilué au $1/20 \times 10^{-4}$ (piste 32) ont été amplifiés par RT-PCR en temps réel à l'aide du kit "Light Cycler RNA Amplification Kit Hybridization Probes" et des couples d'amorces et de sondes de la série n° 2, dans les conditions décrites à l'exemple 8.

- la figure 13 (figure 13.1 à 13.70) représente la carte de restriction
25 de la séquence SEQ ID NO : 1 correspondant à l'équivalent ADN du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589.

- la figure 14 montre le résultat du test de sérologie SRAS par ELISA N indirect (1ère série de sérums testés)

- la figure 15 montre le résultat du test de sérologie SRAS par
30 ELISA N indirect (2ème série de sérums testés)

- la figure 16 présente le résultat du test de sérologie SRAS par ELISA N double épitope (1ère série de sérums testés)

- la figure 17 montre le résultat du test de sérologie SRAS par ELISA N double épitope (2ème série de sérums testés)

5 - la figure 18 illustre le test de réactivité des anticorps monoclonaux anti-N par ELISA sur la nucléoprotéine N native du SRAS-CoV. Les anticorps ont été testés sous la forme de surnageants de culture d'hybridomes par un ELISA indirect utilisant un lysat irradié de cellules VeroE6 infectées par le SRAS-CoV comme antigène (courbes lysat SRAS). Un contrôle négatif de réactivité est réalisé pour
10 chaque anticorps sur un lysat de cellules VeroE6 non infectées (courbes lysat négatif). Plusieurs anticorps monoclonaux de spécificité connue ont été utilisés comme anticorps témoins négatifs : para1-3 dirigé contre les antigènes des virus parainfluenza de type 1-3 (Bio-Rad) et grippe B dirigé contre les antigènes du virus de la grippe de type B (Bio-Rad).

15 - la figure 19 illustre le test de réactivité des anticorps monoclonaux anti-N du SRAS-CoV par ELISA sur les antigènes natifs du coronavirus humain 229E (HCoV-229E). Les anticorps ont été testés sous la forme de surnageants de culture d'hybridomes par un test ELISA indirect utilisant un lysat de cellules MRC-5 infectées par le coronavirus humain 229E comme antigène (courbes lysat 229E). Un
20 contrôle négatif d'immunoréactivité est réalisé pour chaque anticorps sur un lysat de cellules MRC-5 non infectées (courbes lysat négatif). L'anticorps monoclonal 5-11H.6 dirigé contre la protéine S du coronavirus humain 229E (Sizun et al. 1998, J. Virol. Met. 72 : 145-152) est utilisé comme anticorps témoin positif. Les anticorps para1-3 dirigé contre les antigènes des virus parainfluenza de type 1-3 (Bio-Rad) et grippe B
25 dirigé contre les antigènes du virus de la grippe de type B (Bio-Rad) ont été ajoutés au panel des anticorps monoclonaux testés.

- la figure 20 montre un test de réactivité des anticorps monoclonaux anti-N du SRAS-CoV par western blot sur la nucléoprotéine N native du SRAS-CoV dénaturée. Un lysat de cellules VeroE6 infectées par le SRAS-CoV a été préparé dans
30 le tampon de dépôt selon Laemmli et mis à migrer dans un gel SDS à 12% de polyacrylamide puis les protéines ont été transférées sur membrane de PVDF. Les anticorps monoclonaux anti-N testés ont été utilisés pour l'immunoessai à la concen-

tration de 0.05 µg/ml. La révélation est faite avec des anticorps anti-IgG(H+L) de souris couplés à la peroxydase (NA93IV, Amersham) et le système ECL+. Deux anticorps monoclonaux ont été utilisés comme témoins négatifs de réactivité : grippe B dirigé contre les antigènes du virus de la grippe de type B (Bio-Rad) et para1-3 dirigé
5 contre les antigènes des virus parainfluenza de type 1-3 (Bio-Rad).

- la figure 21 présente les plasmides d'expression en cellules de mammifères de la protéine S du SRAS-CoV. Le cDNA de la S du SRAS-CoV a été inséré entre les sites BamH1 et Xho1 du plasmide d'expression pcDNA3.1(+) (Clontech) pour obtenir le plasmide pcDNA-S et entre les sites Nhe1 et Xho1 du
10 plasmide d'expression pCI (Promega) pour obtenir le plasmide pCI-S. Les séquences WPRE et CTE ont été insérées dans chacun des deux plasmides pcDNA-S et pCI-S entre les sites Xho1 et Xba1 pour obtenir respectivement les plasmides pcDNA-S-CTE, pcDNA-S-WPRE, pCI-S-CTE et pCI-S-WPRE.

SP : peptide signal prédit (aa 1-13) avec le logiciel signalP v2.0
15 (Nielsen et al., 1997, Protein Engineering, 10 : 1-6)

TM : région transmembranaire prédite (aa 1196-1218) avec le logiciel TMHMM v2.0 (Sonnhammer et al., 1998, Proc. of Sixth Int. Conf. on Intelligent Systems for Molecular Biology, p 175-182, AAAI Press). Il faut noter que les acides aminés W1194 et P1195 font possiblement partie de la région trans-
20 membranaire avec des probabilités respectives de 0,13 et 0,42

P-CMV : promoteur immédiat/précoce du cytomégalo virus. BGH
pA : signal de polyadénylation du gène de l'hormone de croissance bovine

SV40 late pA : signal de polyadénylation tardif du virus SV40

SD/SA : sites donneur et accepteur d'épissage

25 WPRE : séquences du "Woodchuck Hepatitis Virus posttranscriptional regulatory element" du virus de l'hépatite de la marmotte

CTE : séquences du "constitutive transport element" du rétrovirus simien de Mason-Pfizer

- la figure 22 illustre l'expression de la protéine S après transfection
30 de cellules VeroE6. Des extraits cellulaires ont été préparés 48 heures après transfection de cellules VeroE6 par les plasmides pcDNA, pcDNA-S, pCI et pCI-S. Des extraits cellulaires ont également été préparés 18 heures après infection par le virus

recombinant de la vaccine VV-TF7.3 et transfection par les plasmides pcDNA ou pcDNA-S. A titre de contrôle, des extraits de cellules VeroE6 ont été préparés 8 heures après infection par le SRAS-CoV à une multiplicité d'infection de 3. Ils ont été séparés sur un gel SDS à 8% d'acrylamide et analysés par western blot à l'aide d'un anticorps polyclonal de lapin anti-S et d'un anticorps polyclonal anti-IgG(H+L) de lapin couplé à la peroxidase (NA934V, Amersham). Une échelle de masse moléculaire (kDa) est portée sur la figure.

SRAS-CoV : extrait de cellules VeroE6 infectées par le SRAS-CoV

Mock : extrait contrôle de cellules non infectées

10 - la figure 23 illustre l'effet des séquences CTE et WPRE sur l'expression de la protéine S après transfection de cellules VeroE6 et 293T. Des extraits cellulaires ont été préparés 48 heures après transfection de cellules VeroE6 (A) ou 293T (B) par les plasmides pcDNA, pcDNA-S, pcDNA-S-CTE, pcDNA-S-WPRE, pCI-S, pCI-S-CTE et pCI-S-WPRE, séparés sur un gel SDS à 8% d'acrylamide et
15 analysés par western blot à l'aide d'un anticorps polyclonal de lapin anti-S et d'un anticorps polyclonal anti-IgG(H+L) de lapin couplé à la peroxidase (NA934V, Amersham). Une échelle de masse moléculaire (kDa) est portée sur la figure.

SRAS-CoV : extrait de cellules VeroE6 préparés 8 heures après infection par le SRAS-CoV à une multiplicité d'infection de 3.

20 Mock : extrait contrôle de cellules VeroE6 non infectées

- la figure 24 présente des vecteurs lentiviraux défectifs à DNA flap central pour l'expression de la S du SRAS-CoV. Le cADN de la protéine S du SRAS-CoV a été cloné sous la forme d'un fragment BamH1-Xho1 dans le plasmide pTRIPΔU3-CMV contenant un vecteur lentiviral défectif TRIP à DNA flap central
25 (Sirven et al, 2001, Mol. Ther., 3 :438-448) pour obtenir le plasmide pTRIP-S. Les cassettes d'expression optimales constituées du promoteur immédiat/précoce du virus CMV, d'un signal d'épissage, du cDNA de la S et de l'un ou l'autre des signaux post-transcriptionnels CTE ou WPRE ont été substituées à la cassette EF1α-EGFP du vecteur d'expression lentiviral défectif à DNA FLAP central TRIPΔU3-EF1 α (Sirven
30 et al, 2001, Mol. Ther., 3 :438-448) pour obtenir les plasmides pTRIP-SD/SA-S-CTE et pTRIP-SD/SA-S-WPRE.

SP : peptide signal

TM : région transmembranaire

P-CMV : promoteur immédiat/précoce du cytomégalo virus

P-EF1 α : promoteur du gène EF1 α

SD/SA : sites donneur et accepteur d'épissage

5 WPRE : séquences du "Woodchuck Hepatitis Virus posttranscriptional regulatory element" du virus de l'hépatite de la marmotte

CTE : séquences du "constitutive transport element" du rétrovirus simien de Mason-Pfizer

LTR : « Long terminal repeat »

10 Δ U3 : LTR délété des séquences « promoter/enhancer »

cPPT : « polypurine tract cis-active sequence »

CTS : « central termination sequence »

-la figure 25 montre l'analyse par western blot de l'expression de la S du SRAS-CoV par des lignées cellulaires transduites par les vecteurs lentiviraux
15 TRIP-SD/SA-S-WPRE et TRIP-SD/SA-S-CTE. Des extraits cellulaires ont été préparés à partir de lignées FrhK4-S-CTE et FrhK4-S-WPRE établies après transduction par les vecteurs lentiviraux TRIP-SD/SA-S-CTE et TRIP-SD /SA-S-WPRE respectivement. Ils ont été séparés sur un gel SDS à 8% d'acrylamide et analysés par western blot à l'aide d'un anticorps polyclonal de lapin anti-S et d'un conjugué anti-
20 IgG(H+L) de lapin couplé à la peroxidase. Une échelle de masse moléculaire (kDa) est portée sur la figure.

T- : extrait contrôle de cellules FrhK-4

T+ : extrait de cellules FrhK-4 préparées 24h après infection par le SRAS-CoV à une multiplicité d'infection de 3.

25 - la figure 26 concerne l'analyse de l'expression de polypeptide Ssol par des lignées cellulaires transduites par les vecteurs lentiviraux TRIP-SD/SA-Ssol-WPRE et TRIP-SD/SA-Ssol-CTE. La sécrétion du polypeptide Ssol a été recherchée dans le surnageant d'une série de clones cellulaires isolés après transduction de cellules FRhK-4 par les vecteurs lentiviraux TRIP-SD/SA-Ssol-WPRE et TRIP-
30 SD/SA-Ssol-CTE. 5 μ l de surnageant dilués au 1/2 dans du tampon de dépôt selon Laemmli ont été analysés par western blot révélé par un anticorps monoclonal anti-FLAG (M2, Sigma) et un conjugué anti-IgG(H+L) de souris couplé à la peroxidase.

T- : surnageant de la lignée FRhK-4 parentale. T+ : surnageant de cellules BHK infectées par un virus recombinant de la vaccine exprimant le polypeptide Ssol. La flèche pleine indique le polypeptide Ssol, tandis que la flèche creuse indique une réaction croisée avec une protéine d'origine cellulaire.

5 - la figure 27 montre les résultats relatifs à l'analyse du polypeptide Ssol purifié

A. 8, 2, 0.5 et 0.125 µg de polypeptide recombinant Ssol purifié par chromatographie d'affinité anti-FLAG et gel filtration (G75) ont été séparés sur gel SDS à 8% de polyacrylamide. Le polypeptide Ssol ainsi que des quantités variables de
10 marqueurs de masse moléculaire (MM) ont été révélés par coloration au nitrate d'argent (Gelcode SilverSNAP stain kit II, Pierce).

B. Marqueurs étalons pour l'analyse par spectrométrie de masse SELDI-TOF

IgG : IgG bovine de MM 147300

15 ConA : conalbumine de MM 77490

HRP : peroxydase de raifort analysée à titre de contrôle et de MM
43240

C. Analyse par spectrométrie de masse (SELDI-TOF) du polypeptide recombinant Ssol.

20 Les pics A et B correspondent au polypeptide Ssol simplement et doublement chargé.

D. Séquençage de l'extrémité N-terminale du polypeptide recombinant Ssol. 5 cycles de dégradation d'Edman en phase liquide ont été réalisés sur un séquenceur ABI494 (Applied Biosystems).

25 - la figure 28 illustre l'influence d'un signal d'épissage et des séquences CTE et WPRE sur l'efficacité de l'immunisation génique à l'aide d'ADN plasmidique codant la S du SRAS-CoV

A. Des groupes de 7 souris BALB/c ont été immunisés à deux reprises à 4 semaines d'intervalle à l'aide de 50 µg d'ADN plasmidique de pCI, pcDNA-S, pCI-S, pcDNA-N et pCI-HA.
30

B. Des groupes de 6 souris BALB/c ont été immunisés à deux reprises à 4 semaines d'intervalle à l'aide de 2 µg, 10µg ou 50 µg d'ADN plasmidique de pCI, pCI-S, pCI-S-CTE et pCI-S-WPRE.

Les sérums immuns prélevés 3 semaines après la deuxième immuni-
sation ont été analysés par ELISA indirect en utilisant un lysat de cellules VeroE6 infectées par le SRAS-CoV comme antigène. Les titres en anticorps anti-SRAS-CoV sont calculés comme l'inverse de la dilution produisant une DO spécifique de 0,5 après révélation par un anticorps polyclonal anti-IgG de souris couplé à la peroxidase (NA931V, Amersham) et du TMB (KPL).

- la figure 29 montre la séroneutralisation de l'infectivité du SRAS-CoV par les anticorps induits chez la souris après immunisation génique à l'aide d'ADN plasmidique codant la S du SRAS-CoV. Des pools de sérums immuns prélevés 3 semaines après la deuxième immunisation ont été réalisés pour chacun des groupes des expériences décrites dans la figure 28 et évalués pour leur capacité à séroneutraliser l'infectivité de 100 TCID₅₀ du SRAS-CoV sur cellules FRhK-4. 4 points sont réalisés pour chacune des dilutions de raison 2 testées à partir du 1/20. Le titre séroneutralisant est calculé selon la méthode de Reed et Munsch comme l'inverse de la dilution neutralisant l'infectivité de 2 cupules sur 4.

A. Groupes de souris BALB/c immunisés à deux reprises à 4 semaines d'intervalle à l'aide de 50 µg d'ADN plasmidique de pCI, pcDNA-S, pCI-S, pcDNA-N et pCI-HA. □ : sérum préimmun. ■ : sérum immun.

B. Groupes de souris BALB/c immunisés à deux reprises à 4 semaines d'intervalle à l'aide de 2 µg, 10µg ou 50 µg d'ADN plasmidique de pCI, pCI-S, pCI-S-CTE et pCI-S-WPRE.

- la figure 30 illustre l'immunoréactivité du polypeptide recombinant Ssol vis à vis de sérums de patients atteints de SRAS. La réactivité de sérums de patients a été analysée par test ELISA indirect contre des phases solides préparées à l'aide du polypeptide recombinant Ssol purifié. Les anticorps de patients réagissant avec la phase solide à une dilution de 1/400 sont révélés par un anticorps polyclonal anti-IgG(H+L) humain couplé à la peroxidase (Amersham NA933V) et du TMB plus H₂O₂ (KPL). Les sérums de cas probables de SRAS sont identifiés par un numéro d'ordre du Centre National de Référence des Virus Influenzae ainsi que par les

initiales du patient et le nombre de jours écoulés depuis le début des symptômes, le cas échéant. Les sérums TV sont des sérums témoins de sujets qui ont été prélevés en France avant l'épidémie de SRAS survenue en 2003.

- La figure 31 montre l'induction d'anticorps dirigés contre le SRAS-CoV après immunisation avec le polypeptide recombinant Ssol. Deux groupes de 6 souris ont été immunisés à 3 semaines d'intervalle avec 10 µg de polypeptide recombinant Ssol (groupe Ssol) adjuvé avec de l'hydroxyde d'aluminium ou, à titre de contrôle, de l'adjuvant seul (groupe mock). Trois immunisations successives ont été réalisées et les sérums immuns ont été prélevés 3 semaines après chacune des trois immunisations (IS1, IS2, IS3). Les sérums immuns ont été analysés par pool pour chacun des 2 groupes par ELISA indirect en utilisant un lysat de cellules VeroE6 infectées par le SRAS-CoV comme antigène. Les titres en anticorps anti-SRAS-CoV sont calculés comme l'inverse de la dilution produisant une DO spécifique de 0,5 après révélation par un anticorps polyclonal anti-IgG de souris couplé à la peroxidase (Amersham) et du TMB (KPL).

- La figure 32 présente l'alignement nucléotidique des séquences du gène synthétique 040530 avec la séquence du gène sauvage de l'isolat 031589 du SRAS-CoV. I-3059 correspond aux nucléotides 21406-25348 de l'isolat 031589 du SRAS-CoV déposé à la C.N.C.M. sous le numéro I-3059 (SEQ ID NO : 4, plasmide pSRAS-S). S-040530 est la séquence du gène synthétique 040530.

- la figure 33 illustre l'utilisation d'un gène synthétique pour l'expression de la S du SRAS-CoV. Des extraits cellulaires préparés 48 heures après transfection de cellules VeroE6 (A) ou 293T (B) par les plasmides pCI, pCI-S, pCI-S-CTE, pCI-S-WPRE et pCI-Ssynth ont été séparés sur un gel SDS à 8% d'acrylamide et analysés par western blot à l'aide d'un anticorps polyclonal de lapin anti-S et d'un anticorps polyclonal anti-IgG(H+L) de lapin couplé à la peroxidase (NA934V, Amersham). Le western blot est révélé par luminescence (ECL+, Amersham) et acquisition sur un dispositif d'imagerie numérique (FluorS, BioRad). Les niveaux d'expression de la protéine S ont été mesurés en quantifiant les 2 bandes majoritaires repérées sur l'image.

- la figure 34 présente un schéma de la construction des virus vaccins recombinants VV-TG-S, VV-TG-Ssol, VV-TN-S et VV-TN-Ssol

A. Les cDNA de la protéine S et du polypeptide Ssol du SRAS-CoV ont été insérés entre les sites BamH1 et Sma1 du plasmide de transfert pTG186 pour obtenir les plasmides pTG-S et pTG-Ssol.

5 B. Les séquences du promoteur synthétique 480 ont ensuite été substituées à celles du promoteur 7.5 par échange du fragment Nde1-Pst1 des plasmides pTG186poly, pTG-S et pTG-Ssol pour obtenir les plasmides de transfert pTN480, pTN-S et pTN-Ssol.

C. Séquence du promoteur synthétique 480 tel que contenu entre les sites Nde1 et Pst1 des plasmides de transfert de la série pTN. Un site Asc1 a été inséré
10 pour faciliter les manipulations ultérieures. Les sites de restriction ainsi que la séquence du promoteur sont soulignés

D. Les virus recombinants de la vaccine sont obtenus par double recombinaison homologue *in vivo* entre la cassette TK des plasmides de transfert des séries pTG et pTN et le gène TK de la souche Copenhague du virus de la vaccine.

15 SP : peptide signal prédit (aa 1-13) avec le logiciel signalP v2.0 (Nielsen et al., 1997, Protein Engineering, 10 : 1-6)

TM : région transmembranaire prédite (aa 1196-1218) avec le logiciel TMHMM v2.0 (Sonnhammer et al., 1998, Proc. of Sixth Int. Conf. on Intelligent Systems for Molecular Biology, p 175-182, AAAI Press). Il faut noter que
20 les acides aminés W1194 et P1195 font possiblement partie de la région transmembranaire avec des probabilités respectives de 0,13 et 0,42.

TK-L, TK-R : parties gauche et droite du gène de la thymidine kinase du virus de la vaccine

MCS : site multiple de clonage

25 PE : promoteur précoce

PL : promoteur tardif

PL synth : promoteur tardif synthétique 480

- la figure 35 illustre l'expression de la protéine S par les virus vaccine recombinants, analysée par western blot. Des extraits cellulaires ont été préparés 18 heures après infection de cellules CV1 par les virus vaccine recombinants VV-TG, VV-TG-S et VV-TN-S à une M.O.I. de 2 (A). A titre de contrôle, des extraits de
30 cellules VeroE6 ont été préparés 8 heures après infection par le SRAS-CoV à une

multiplicité d'infection de 2. Des extraits cellulaires ont également été préparés 18 heures après infection de cellules CV1 par les virus vaccine recombinants VV-TG-S, VV-TG-Ssol, VV-TN, VV-TN-S et VV-TN-Ssol (B). Ils ont été séparés sur des gels SDS à 8% d'acrylamide et analysés par western blot à l'aide d'un anticorps polyclonal de lapin anti-S et d'un anticorps polyclonal anti-IgG(H+L) de lapin couplé à la peroxidase (NA934V, Amersham). « 1µl » et « 10µl » indique les quantités d'extraits cellulaires déposées sur le gel. Une échelle de masse moléculaire (kDa) est portée sur la figure.

SRAS-CoV : extrait de cellules VeroE6 infectées par le SRAS-CoV
Mock : extrait contrôle de cellules non infectées
- la figure 36 montre le résultat d'une analyse par western blot de la sécrétion du polypeptide Ssol par les virus vaccine recombinants.

A. Des surnageants de cellules CV1 infectées par le virus vaccine recombinant VV-TN, différents clones du virus VV-TN-Ssol et par les virus VV-TG-Ssol ou VV-TN-Sflag ont été récoltés 18 heures après infection de cellules CV1 à une M.O.I. de 2.

B. Des surnageants de cellules 293T, FRhK-4, BHK-21 et CV1 infectées en dupliqués (1,2) par le virus vaccine recombinant VV-TN-Ssol à une M.O.I. de 2 ont été récoltés 18 heures après infection. Le surnageant de cellules CV1 infectées par le virus VV-TN a également été récolté à titre de contrôle (M).

Tous les surnageants ont été séparés sur un gel SDS à 8% d'acrylamide selon Laemmli et analysés par western blot à l'aide d'un anticorps monoclonal de souris anti-FLAG et d'un anticorps polyclonal anti-IgG(H+L) de souris couplé à la peroxidase (NA931V, Amersham) (A) ou à l'aide d'un anticorps polyclonal de lapin anti-S et d'un anticorps polyclonal anti-IgG(H+L) de lapin couplé à la peroxidase (NA934V, Amersham) (B).

Une échelle de masse moléculaire (kDa) est portée sur la figure.

- la figure 37 montre l'analyse du polypeptide Ssol, purifié par gel SDS de polyacrylamide

10, 5 et 2 µl de polypeptide recombinant Ssol purifié par chromatographie d'affinité anti-FLAG ont été séparés sur gel SDS en gradient de 4 à 15 % de polyacrylamide. Le polypeptide Ssol ainsi que des quantités variables de marqueurs de

masse moléculaire (MM) ont été révélés par coloration au nitrate d'argent (Gelcode SilverSNAP stain kit II, Pierce).

- la figure 38 illustre l'immunoréactivité du polypeptide recombinant Ssol produit par le virus vaccine recombinant VV-TN-Ssol vis-à-vis de sérums de patients atteints de SRAS. La réactivité de sérums de patients a été analysée par test ELISA indirect contre des phases solides préparées à l'aide du polypeptide recombinant Ssol purifié. Les anticorps de patients réagissant avec la phase solide à une dilution de 1/100 et 1/400 sont révélés par un anticorps polyclonal anti-IgG(H+L) humain couplé à la peroxidase (Amersham NA933V) et du TMB plus H₂O₂ (KPL). Les sérums de cas probables de SRAS sont identifiés par un numéro d'ordre du Centre National de Référence des Virus Influenzae ainsi que par les initiales du patient et le nombre de jours écoulés depuis le début des symptômes, le cas échéant. Les sérums TV sont des sérums témoins de sujets qui ont été prélevés en France avant l'épidémie de SRAS survenue en 2003.

- la figure 39 montre la réponse en anticorps anti-SRAS-CoV chez la souris après immunisation par les virus vaccine recombinants. Des groupes de 7 souris BALB/c ont été immunisés par voie i.v. à deux reprises à 4 semaines d'intervalle par 106 u.f.p. de virus vaccine recombinants VV-TG, VV-TG-HA, VV-TG-S, VV-TG-Ssol, VV-TN, VV-TN-S, VV-TN-Ssol.

A. Des pools de sérums immuns prélevés 3 semaines après chacune des deux immunisations ont été réalisés pour chacun des groupes et ont été analysés par ELISA indirect en utilisant un lysat de cellules VeroE6 infectées par le SRAS-CoV comme antigène. Les titres en anticorps anti-SRAS-CoV sont calculés comme l'inverse de la dilution produisant une DO spécifique de 0,5 après révélation par un anticorps polyclonal anti-IgG de souris couplé à la peroxidase (NA931V, Amersham) et du TMB (KPL).

B. Les pools de sérums immuns ont été évalués pour leur capacité à séroneutraliser l'infectivité de 100 TCID₅₀ du SRAS-CoV sur cellules FRhK-4. 4 points sont réalisés pour chacune des dilutions de raison 2 testées à partir du 1/20. Le titre séroneutralisant est calculé selon la méthode de Reed et Munsch comme l'inverse de la dilution neutralisant l'infectivité de 2 cupules sur 4.

- la figure 40 décrit la construction des virus recombinants MVSchw2-SARS-S et MVSchw2-SARS-Ssol.

A. Le vecteur rougeole est un génome complet de la souche vaccinale Schwarz du virus de la rougeole (MV) dans lequel une unité de transcription supplémentaire a été introduite (Combredet, 2003, Journal of Virology, 77 : 11546-11554). L'expression des phases ouvertes de lecture (ORF) supplémentaires est contrôlée par les éléments agissant en cis nécessaires à la transcription, à la formation de la coiffe et à la polyadénylation du transgène, qui ont été copiés des éléments présents à la jonction N/P. 2 vecteurs différents permettent l'insertion entre les gènes P (phosphoprotéine) et M (matrice) d'une part et H (hémagglutinine) et L (polymérase) d'autre part.

B. Les génomes recombinants MVSchw2-SARS-S et MVSchw2-SARS-Ssol du virus de la rougeole ont été construits par l'insertion des ORFs de la protéine S et du polypeptide Ssol au sein d'une unité de transcription supplémentaire localisée entre les gènes P et M du vecteur.

Les différents gènes du virus de la rougeole (MV) sont indiqués : N (nucléoprotéine), PVC (phosphoprotéine et protéine V/C), M (matrice), F (fusion), H (hémagglutinine), L (polymérase). T7 = promoteur de l'ARN polymérase T7, hh = ribozyme hammerhead, T7t = séquence terminatrice de l'ARN polymérase du phage T7, ∂ = ribozyme du virus de l'hépatite ∂ , (2), (3) = unités de transcription supplémentaires (ATU).

Taille du génome du MV : 15894 nt.

SP : peptide signal

TM : région transmembranaire

25 FLAG : étiquette FLAG

- la figure 41 illustre l'expression de la protéine S par les virus rougeole recombinants, analysée par western blot.

Des extraits cytoplasmiques ont été préparés après infection de cellules Vero par différents passages des virus MVSchw2-SARS-S et MVSchw2-SARS-Ssol et le virus sauvage MWSchw à titre de contrôle. Des extraits cellulaires en tampon de dépôt selon Laemmli ont également été préparés 8 heures après infection de cellules VeroE6 par le SRAS-CoV à une multiplicité d'infection de 3. Ils ont été

séparés sur un gel SDS à 8% d'acrylamide et analysés par western blot à l'aide d'un anticorps polyclonal de lapin anti-S et d'un anticorps polyclonal anti-IgG(H+L) de lapin couplé à la peroxidase (NA934V, Amersham).

Une échelle de masse moléculaire (kDa) est portée sur la figure.

5 Pn : nième passage du virus après coculture de cellules 293-3-46 et Vero.

SRAS-CoV : extrait de cellules VeroE6 infectées par le SRAS-CoV

Mock : extrait contrôle de cellules VeroE6 non infectées

- la figure 42 montre l'expression de la protéine S par les virus
10 rougeole recombinants, analysée par immunofluorescence

Des cellules Vero en monocouches sur lamelles de verre ont été infectées par le virus sauvage MWSchw (A) ou les virus MVSchw2-SARS-S (B) et MVSchw2-SARS-Ssol (C). Quand les syncytia ont atteint 30 à 40% de confluence (A., B.) ou 90-100% (C), les cellules ont été fixées, perméabilisées et marquées par
15 des anticorps polyclonaux de lapins anti-SRAS-CoV et un conjugué anti-IgG(H+L) de lapin couplé au FITC (Jackson).

- la figure 43 illustre l'analyse par western blot de l'immunoréactivité de sérums de lapins dirigés contre les peptides E1-12, E53-76 et M2-14. Le lapin 20047 a été immunisé avec le peptide E1-12 couplé à la KLH. Les
20 lapins 22234 et 22240 ont été immunisés avec le peptide E53-76 couplé à la KLH. Les lapins 20013 et 20080 ont été immunisés avec le peptide M2-14 couplé à la KLH. Les immunsérums ont été analysés par western blot à l'aide d'extraits de cellules infectées par le SRAS-CoV (B) ou à l'aide d'extraits de cellules infectées par un virus recombinant de la vaccine exprimant la protéine E (A) ou M (C) de l'isolat 031589 du SRAS-
25 CoV. Les immunoempreintes ont été révélées à l'aide d'un conjugué anti-IgG(H+L) de lapin couplé à la peroxidase (NA934V, Amersham).

La position des protéines E et M est indiquée par une flèche.

Une échelle de masse moléculaire (kDa) est portée sur la figure.

Il doit être bien entendu, toutefois, que ces exemples sont donnés
30 uniquement à titre d'illustration de l'objet de l'invention, dont ils ne constituent en aucune manière une limitation.

Exemple 1 : Clonage et séquençage du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589

L'ARN de la souche de SARS-CoV a été extrait à partir du prélèvement de lavage bronchoalvéolaire répertorié sous le numéro 031589, effectué sur un patient de l'hôpital français de Hanoi (Vietnam) atteint de SRAS.

L'ARN isolé a été utilisé comme matrice pour amplifier les ADNc correspondant aux différents cadres ouverts de lecture du génome (ORF 1a, ORF1b, ORF-S, ORF-E, ORF-M, ORF-N (incluant les ORF-13 et ORF-14), ORF3, ORF4, ORF7 à ORF11), et aux extrémités 5' et 3' non-codantes. Les séquences des amorces et des sondes utilisées pour l'amplification/détection ont été définies d'après la séquence nucléotidique disponible du SARS-CoV.

Dans ce qui suit les amorces et les sondes sont identifiées par : la lettre S, suivie d'une lettre qui indique la région correspondante du génome (L pour l'extrémité 5' incluant ORF1a et ORF1b ; S, M et N pour les ORF-S, ORF-M, ORF-N, SE et MN pour les régions intergéniques correspondantes), puis éventuellement de Fn, Rn, avec n inclus entre 1 et 6 correspondant aux amorces utilisées pour la PCR nichée ou imbriquée (paire F1 + R1 pour la première amplification, paire F2 + R2 pour la deuxième amplification, etc...), puis de +/- ou -/- correspondant à une amorce sens ou antisens et enfin des positions des amorces en référence à la séquence Genbank AY27411.3 ; pour les amorces S et N sens et antisens et les autres amorces sens uniquement, lorsqu'une seule position est indiquée elle correspond à celle de l'extrémité 5' d'une sonde ou d'une amorce d'environ 20 bases ; pour les amorces antisens autres que les amorces S et N, lorsqu'une seule position est indiquée elle correspond à celle de l'extrémité 3' d'une sonde ou d'une amorce d'environ 20 bases.

Les produits d'amplifications ainsi générés ont été séquencés à l'aide d'amorces spécifiques afin de déterminer la séquence complète du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589. Ces produits d'amplification, à l'exception de ceux correspondant aux ORF1a et ORF1b, ont ensuite été clonés dans des vecteurs d'expression afin de produire les protéines virales correspondantes et les anticorps dirigés contre ces protéines, notamment par immunisation à base d'ADN.

1. Extraction des ARN

Les ARN ont été extraits à l'aide du kit *QIamp viral RNA extraction mini* (QIAGEN) en suivant les recommandations du fabricant. De manière plus précise : 140 µl du prélèvement et 560 µl de tampon AVL ont été mélangés vigoureu-
5 sement pendant 15 secondes, incubés 10 min à température ambiante puis centrifugés brièvement à vitesse maximale. 560 µl d'éthanol à 100% ont été ajoutés au surnageant et le mélange ainsi obtenu a été agité très vigoureusement pendant 15 sec. 630 µl du mélange ont ensuite été déposés sur la colonne.

La colonne a été placée sur un tube de 2 ml, centrifugée 1 min à
10 8000 rpm, puis le reste du mélange précédent a été déposé sur la même colonne, centrifugé à nouveau, 1 min à 8000 rpm et la colonne a été transférée sur un tube de 2 ml propre. Ensuite, 500 µl de tampon AW1 ont été ajoutés sur la colonne, puis la colonne a été centrifugée 1 min à 8000 rpm et l'éluat a été éliminé. 500 µl de tampon AW2 ont été ajoutés sur la colonne qui a ensuite été centrifugée 3 min à 14000 rpm et
15 transférée sur un tube de 1,5 ml. Enfin, 60 µl de tampon AVE ont été ajoutés sur la colonne qui a été incubée 1 à 2 min à température ambiante puis centrifugée 1 min à 8000 rpm. L'éluat correspondant à l'ARN purifié a été récupéré et congelé à -20°C.

2. Amplification, séquençage et clonage des ADNc

2.1) ADNc codant pour la protéine S

20 Les ARN extraits à partir du prélèvement ont été soumis à une transcription inverse à l'aide d'oligonucléotides hexamériques de séquence aléatoire (pdN6), afin de produire des fragments d'ADNc.

La séquence codant pour la glycoprotéine S du SARS-CoV a été amplifiée sous la forme de deux fragments d'ADN chevauchants : fragment 5'
25 (SRAS-Sa, SEQ ID NO:5) et fragment 3' (SRAS-Sb, SEQ ID NO:6), en réalisant deux amplifications successives à l'aide d'amorces imbriquées. Les amplicons ainsi obtenus ont été séquencés, clonés dans le vecteur plasmidique PCR 2.1-TOPO™ (IN VITROGEN), puis la séquence des ADNc clonés a été déterminée.

a) clonage et séquençage des fragments Sa et Sb

30 a₁) synthèse de l'ADNc

Le mélange réactionnel contenant : ARN (5 µl) , H₂O ppi (3,5 µl),
tampon de transcriptase inverse5X (4 µl), dNTP 5 mM (2 µl), pdN6 100 ug/ml (4 µl),

RNasin 40 UI/ul (0,5 µl) et transcriptase inverse AMV-RT, 10 UI/ul, PROMEGA (1µl) a été incubé dans un thermocycleur dans les conditions suivantes : 45 min à 42°C, 15 min à 55°C, 5 min à 95°C, puis l'ADNc obtenu a été maintenu à +4°C.

a₂) première amplification PCR

5 Les extrémités 5' et 3' du gène S ont été amplifiées respectivement avec les paires d'amorces S/F1/+ 21350-21372 et S/R1/- 23518-23498, S/F3/+ 23258-23277 et S/R3/- 25382-25363. Le mélange réactionnel de 50 µl contenant : ADNc (2 µl), amorces 50 µM (0,5 µl), tampon 10 X (5 µl), dNTP 5 mM (2 µl), Taq Expand High Fidelity, Roche (0,75 µl) et H₂O (39, 75 µl) a été amplifié dans un
10 thermocycleur, dans les conditions suivantes : une étape initiale de dénaturation à 94°C pendant 2 min a été suivie de 40 cycles comprenant : une étape de dénaturation à 94°C pendant 30 sec, une étape d'hybridation à 55°C pendant 30 sec puis une étape d'élongation à 72°C pendant 2 min 30 sec, avec 10 sec d'élongation supplémentaire à chaque cycle, puis d'une étape finale d'élongation à 72°C pendant 5 min.

15 a₃) deuxième amplification PCR

Les produits de la première amplification PCR (amplicons 5' et 3') ont subi une seconde étape d'amplification PCR (PCR nichée) dans des conditions identiques à celles de la première amplification, avec les paires d'amorces S/F2/+ 21406-21426 et S/R2/- 23454-23435, et S/F4/+ 23322-23341 et S/R4/- 25348-
20 25329, respectivement pour l'amplicon 5' et l'amplicon 3'.

a₄) clonage et séquençage des fragments Sa et Sb

Les amplicons Sa (extrémité 5') et Sb (extrémité 3') ainsi obtenus ont été purifiés à l'aide du kit *QIAquick PCR purification* (QIAGEN), en suivant les recommandations du fabricant, puis ils ont été clonés dans le vecteur PCR2.1-TOPO
25 (kit Invitrogen), pour donner les plasmides dénommés SRAS-S1 et SRAS-S2.

L'ADN des clones Sa et Sb a été isolé puis l'insert correspondant a été séquencé à l'aide du Kit Big Dye, Applied Biosystem® et des amorces universelles M13 forward et M13 reverse, ainsi que des amorces : S/S/+ 21867, S/S/+ 22353, S/S/+ 22811, S/S/+ 23754, S/S/+ 24207, S/S/+ 24699, S/S/+ 24348,
30 S/S/- 24209, S/S/- 23630, S/S/- 23038, S/S/- 22454, S/S/- 21815, S/S/- 24784, S/S/+ 21556, S/S/+ 23130 et S/S/+ 24465, en suivant les instructions du fabricant ; les

séquences des fragments Sa et Sb ainsi obtenues correspondent aux séquences SEQ ID NO :5 et SEQ ID NO :6 dans la liste de séquences jointe en annexe.

Le plasmide, dénommé SARS-S1 a été déposé sous le n° I-3020, le 12 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient un fragment 5' de la séquence du gène S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, lequel fragment dénommé Sa correspondant aux nucléotides des positions 21406 à 23454 (SEQ ID NO :5), en référence à la séquence Genbank AY274119.3 Tor2.

10 Le plasmide, dénommé TOP10F'-SARS-S2 a été déposé sous le n° I-3019, le 12 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient un fragment 3' de la séquence du gène S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, lequel fragment
15 dénommé Sb correspondant aux nucléotides des positions 23322 à 25348 (SEQ ID NO : 6), en référence à la séquence Genbank n° d'accès AY274119.3.

b) clonage et séquençage de l'ADNc complet (clone SRAS-S de 4 kb)

L'ADNc S complet a été obtenu à partir des clones SARS-S1 et SARS-S2 précités, de la façon suivante :

20 1) une réaction d'amplification PCR a été réalisée sur un clone SARS-S2 en présence de l'amorce S/R4/-/25348-25329 précitée et de l'amorce S/S+/24696-24715: un amplicon de 633 pb a été obtenu,

2) une autre réaction d'amplification PCR a été réalisée sur un autre clone SARS-S2, en présence des amorces S/F4+/23322-23341 précitée et S/S/-
25 /24803-24784: un amplicon de 1481 pb a été obtenu,

La réaction d'amplification a été réalisée dans les conditions telles que définies ci-dessus pour l'amplification des fragments Sa et Sb, à l'exception que 30 cycles d'amplifications comprenant une étape de dénaturation à 94° C pendant 20 sec et une étape d'élongation à 72° C pendant 2 min 30 sec ont été effectués.

30 3) les 2 amplicons (633 pb et 1481 pb) ont été purifiés dans les conditions telles que définies ci-dessus pour les fragments Sa et Sb.

4) une autre réaction d'amplification PCR à l'aide des amorces S/F4+/23322-23341 et S/R4-/25348-25329 précitées, a été réalisée sur les amplicons purifiés obtenus en 3). La réaction d'amplification a été réalisée dans les conditions telles que définies ci-dessus pour l'amplification des fragments Sa et Sb, à l'exception
5 que 30 cycles d'amplifications ont été effectués.

L'amplicon de 2026 pb ainsi obtenu a été purifié, cloné dans le vecteur PCR2.1-TOPO puis séquencé comme ci-dessus, à l'aide des amorces telles que définies ci-dessus pour les fragments Sa et Sb. Le clone ainsi obtenu a été dénommé clone 3'.

10 5) Le clone SARS-S1 précédemment obtenu et le clone 3' ont été digérés par *EcoR I*, les bandes d'environ 2kb ainsi obtenues ont été purifiées sur gel puis amplifiées par PCR avec les amorces S/F2+/21406-21426 et S/R4-/25348-25329 précitées. La réaction d'amplification a été réalisée dans les conditions telles que définies ci-dessus pour l'amplification des fragments Sa et Sb, à l'exception que
15 30 cycles d'amplifications ont été effectués. L'amplicon d'environ 4 kb a été purifié et séquencé. Il a ensuite été cloné dans le vecteur PCR2.1-TOPO pour donner le plasmide, dénommé SARS-S, et l'insert contenu dans ce plasmide a été séquencé comme ci-dessus, à l'aide des amorces telles que définies ci-dessus pour les fragments Sa et Sb. Les séquences d'ADNc de l'insert et de l'amplicon codant pour la protéine
20 S, correspondent respectivement aux séquences SEQ ID NO : 4 et SEQ ID NO : 2 dans la liste de séquences jointe en annexe, elles codent pour la protéine S (SEQ ID NO : 3).

La séquence de l'amplicon correspondant à l'ADNc codant pour la protéine S de la souche de SARS-CoV issue du prélèvement n°031589 présente les
25 deux mutations suivantes par rapport aux séquences correspondantes de respectivement les isolats Tor2 et Urbani, les positions des mutations étant indiquées en référence à la séquence complète du génome de l'isolat Tor2 (Genbank AY274119.3) :

- g/t en position 23220 ; le codon alanine (gct) en position 577 de la séquence en acides aminés de la protéine S de Tor2 est remplacé par un codon sérine
30 (tct),

- c/t en position 24872 : cette mutation ne modifie pas la séquence en acides aminés de la protéine S, et

Le plasmide, dénommé SARS-S, a été déposé sous le n° I-3059, le 20 juin 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc codant pour la protéine S de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, laquelle séquence correspondant aux nucléotides des positions 21406 à 25348 (SEQ ID NO :4), en référence à la séquence Genbank AY274119.3.

2.2) ADNc codant pour les protéines M et E

Les ARN issus du prélèvement 031589, extraits comme ci-dessus, ont été soumis à une transcription inverse, associée, lors de la même étape (kit *Titan One Step RT-PCR*®, Roche), à une réaction d'amplification par PCR, à l'aide des couples d'amorces :

- S/E/F1/+ /26051-26070 et S/E/R1/- /26455-26436 pour amplifier l'ORF-E, et
- S/M/F1/+ /26225-26244 et S/M/R1/- /27148-27129 pour amplifier l'ORF-M.

Un premier mélange réactionnel contenant : 8,6 µl d'H₂Oppi, 1 µl de dNTP (5mM), 0,2 µl de chacune des amorces (50µM), 1,25 µl de DTT (100mM) et 0,25 µl de RNAsin (40UI/µl) a été combiné avec un deuxième mélange réactionnel contenant : 1 µl d'ARN, 7 µl d'H₂Oppi, 5 µl de tampon de RT-PCR 5X et 0,5 µl de mélange d'enzyme et les mélanges combinés ont été incubés dans un thermocycleur dans les conditions suivantes : 30 min à 42°C, 10 min à 55°C, 2 min à 94°C suivi de 40 cycles comprenant une étape de dénaturation à 94°C pendant 10 sec, une étape d'hybridation à 55°C pendant 30 sec et une étape d'élongation à 68°C pendant 45 sec, avec 3 sec d'incrément par cycle et enfin une étape d'élongation terminale à 68°C pendant 7 min.

Les produits d'amplification ainsi obtenus (amplicons M et E) ont subi une deuxième amplification PCR (PCR nichée) en utilisant le kit *Expand High-Fi*®, Roche), à l'aide des couples d'amorces :

- S/E/F2/+ /26082-26101 et S/E/R2/- /26413-26394 pour l'amplicon E, et
- S/M/F2/+ /26330-26350 et S/M/R2/- /27098-27078 pour l'amplicon M.

Le mélange réactionnel contenant : 2 µl du produit de la première PCR, 39,25 µl d'H₂Oppi, 5 µl de tampon 10X contenant du MgCl₂, 2 µl de dNTP (5mM), 0,5µl de chacune des amorces (50 µM) et 0,75µl de mélange d'enzyme a été incubé dans un thermocycleur dans les conditions suivantes : une étape de dénaturation à 94°C

pendant 2 min a été suivie de 30 cycles comprenant une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 60°C pendant 30 sec et une étape d'élongation à 72°C pendant 45 sec, avec 3 sec d'incrément par cycle, et enfin une étape d'élongation terminale à 72°C pendant 7 min. Les produits d'amplification
5 obtenus correspondant aux ADNc codant pour les protéines E et M ont été séquencés comme ci-dessus, à l'aide des amorces : S/E/F2+/26082 et S/E/R2-/26394, S/M/F2+/26330, S/M/R2-/27078 précitées et des amorces S/M+/26636-26655 et S/M-/26567-26548. Ils ont ensuite été clonés, comme ci-dessus, pour donner les plasmides dénommés SARS-E et SARS-M. L'ADN de ces clones a ensuite été isolé et
10 séquencé à l'aide des amorces universelles M13 forward et M13 reverse ainsi que des amorces S/M+/26636 et S/M-/26548 précitées.

La séquence de l'amplicon représentant l'ADNc codant pour la protéine E (SEQ ID NO : 13) de la souche de SARS-CoV issue du prélèvement n°031589 ne comporte pas de différences par rapport aux séquences correspondantes
15 des isolats AY274119.3-Tor2 et AY278741-Urbani. La séquence de la protéine E de la souche de SARS-CoV 031589 correspond à la séquence SEQ ID NO : 14 dans la liste de séquences jointe en annexe.

Le plasmide, dénommé SARS-E a été déposé sous le n° I-3046, le 28 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25
20 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc codant pour la protéine E de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 26082 à 26413 (SEQ ID NO :15), en référence à la séquence Genbank n° d'accès AY274119.3.

25 La séquence de l'amplicon représentant l'ADNc codant pour la M (SEQ ID NO :16) de la souche de SARS-CoV issue du prélèvement n°031589 ne comporte pas de différences par rapport à la séquence correspondante de l'isolat AY274119.3-Tor2. En revanche, en position 26857, l'isolat AY278741-Urbani comporte un c et la séquence de la souche de SARS-CoV issue du prélèvement réper-
30 torié sous le n°031589 un t. Cette mutation aboutit à une modification de la séquence en acides aminés de la protéine correspondante: en position 154, une proline (AY278741-Urbani) est changée en sérine dans la souche de SARS-CoV issue du

prélèvement répertorié sous le n°031589. La séquence de la protéine M de la souche de SARS-CoV issue du prélèvement répertorié sous le n°031589 correspond à la séquence SEQ ID NO :17 dans la liste de séquences jointe en annexe.

Le plasmide, dénommé SARS-M a été déposé sous le n° I-3047, le 28 mai 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc codant pour la protéine M de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus ; laquelle séquence correspondant aux nucléotides des positions 26330 à 27098 (SEQ ID NO :18), en référence à la séquence Genbank n° d'accès AY274119.3.

2.3) ADNc correspondant aux ORF3, ORF4, ORF7 à ORF11

La même stratégie d'amplification, de clonage et de séquençage a été utilisée pour obtenir les fragments d'ADNc correspondant respectivement aux ORF suivantes : ORF 3, ORF4, ORF7, ORF8, ORF9, ORF10 et ORF11. Les couples d'amorces utilisées pour la première amplification sont :

- ORF3 et ORF4 : S/SE/F1/+ /25069-25088 et S/SE/R1/- /26300-26281
- ORF7 à ORF11 : S/MN/F1/+ /26898-26917 et S/MN/R1/- /28287-28266

Les couples d'amorces utilisées pour la deuxième amplification sont :

- ORF3 et ORF4 : S/SE/F2/+ /25110-25129 et S/SE/R2/- /26244-26225
- ORF7 à ORF11 : S/MN/F2/+ /26977-26996 et S/MN/R2/- /28218-28199

Les conditions de la première amplification (RT-PCR) sont les suivantes : 45 min à 42°C, 10 min à 55°C, 2 min à 94°C suivi de 40 cycles comprenant une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 58°C pendant 30 sec et une étape d'élongation à 68°C pendant 1 min, avec 5 sec d'incrément par cycle et enfin une étape d'élongation terminale à 68°C pendant 7 min.

Les conditions de la PCR nichée sont les suivantes : une étape de dénaturation à 94°C pendant 2 min a été suivie de 40 cycles comprenant une étape de dénaturation à 94°C pendant 20 sec, une étape d'hybridation à 58°C pendant 30 sec et une étape d'élongation à 72°C pendant 50 sec, avec 4 sec d'incrément par cycle et enfin une étape d'élongation terminale à 72°C pendant 7 min.

Les produits d'amplification obtenus correspondant aux ADNc contenant respectivement les ORF3 et 4 et les ORF7 à 11 ont été séquencés à l'aide

des amorces : S/SE+/25363, S/SE+/25835, S/SE-/25494, S/SE-/25875, S/MN+/27839, S/MN+/27409, S/MN-/27836 S/MN-/27799 et clonés comme ci-dessus pour les autres ORF, pour donner les plasmides dénommés SARS-SE et SARS-MN. L'ADN de ces clones a été isolé et séquencé à l'aide de ces mêmes
5 amorces et des amorces universelles M13 sens et M13 anti-sens.

La séquence de l'amplicon représentant l'ADNc de la région contenant les ORF 3 et 4 (SEQ ID NO :7) de la souche de SARS-CoV issue du prélèvement n°031589 comporte une différence nucléotidique par rapport à la séquence correspondante de l'isolat AY274119-Tor2. Cette mutation en position 25298 aboutit à une
10 modification de la séquence en acides aminés de la protéine correspondante (ORF 3): en position 11, une arginine (AY274119-Tor2) est changée en glycine dans la souche de SARS-CoV issue du prélèvement n°031589. En revanche, aucune mutation n'a été identifiée par rapport à la séquence correspondante de l'isolat AY278741-Urbani. Les séquences des ORF 3 et 4 la souche de SARS-CoV issue du prélèvement n°031589
15 correspondent respectivement aux séquences SEQ ID NO :10 et 12 dans la liste de séquences jointe en annexe.

Le plasmide, dénommé SARS-SE a été déposé sous le n° I-3126, le
13 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient
20 l'ADNc correspondant à la région située entre l'ORF-S et l'ORF-E et chevauchant l'ORF-E de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle région correspondant aux nucléotides des positions 25110 à 26244 (SEQ ID NO :8), en référence à la séquence Genbank n° d'accès AY274119.3,

25 La séquence de l'amplicon représentant l'ADNc correspondant à la région contenant les ORF7 à ORF11 (SEQ ID NO :19) de la souche de SARS-CoV issue du prélèvement n°031589 ne comporte pas de différences par rapport aux séquences correspondantes des isolats AY274119-Tor2 et AY278741-Urbani. Les séquences des ORF7 à 11 de la souche de SARS-CoV issue du prélèvement n°031589
30 correspondent respectivement aux séquences SEQ ID NO : 22, 24, 26, 28 et 30 dans la liste de séquences jointe en annexe.

Le plasmide dénommé SARS-MN a été déposé sous le n° I-3125, le 13 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient la séquence d'ADNc correspondant à la région située entre l'ORF-M et l'ORF-N de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589 et prélevée à Hanoi, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 26977 à 28218 (SEQ ID NO :20), en référence à la séquence Genbank n° d'accès AY274119.3.

La séquence de l'amplicon représentant l'ADNc correspondant à la région contenant les ORF7 à ORF11 (SEQ ID NO :19) de la souche de SARS-CoV issue du prélèvement n°031589 ne comporte pas de différences par rapport aux séquences correspondantes des isolats AY274119-Tor2 et AY278741-Urbani. Les séquences des ORF7 à 11 de la souche de SARS-CoV issue du prélèvement n°031589 correspondent respectivement aux séquences SEQ ID NO : 22, 24, 26, 28 et 30 dans la liste de séquences jointe en annexe.

2.4) ADNc codant pour la protéine N et incluant les ORF13 et ORF14

L'ADNc a été synthétisé et amplifié comme décrit ci-dessus pour les fragments Sa et Sb. De manière plus précise, le mélange réactionnel contenant : 5 µl d'ARN, 5 µl d'H₂O ppi 4 µl de tampon de reverse transcriptase 5X, 2 µl de dNTP (5 mM), 2 µl d'oligo 20T (5 µM), 0,5 µl de RNasin (40 UI/ul) et 1, 5 µl de AMV-RT (10 UI/ul Promega) a été incubé dans un thermocycleur dans les conditions suivantes : 45 min à 42°C, 15 min à 55°C, 5 min à 95°C, puis il a été maintenu à +4°C.

Une première amplification PCR a été réalisée avec la paire d'amorces S/N/F3/+ /28023 et S/N/R3/- /29480.

Le mélange réactionnel comme ci-dessus pour l'amplification des fragments S1 et S2 a été incubé dans un thermocycleur, dans les conditions suivantes : une étape initiale de dénaturation à 94°C pendant 2 min a été suivie de 40 cycles comprenant une étape de dénaturation à 94°C pendant 20 sec, une étape d'hybridation à 55°C pendant 30 sec puis une étape d'élongation à 72°C pendant 1 min 30 sec avec 10 sec d'élongation supplémentaire à chaque cycle, puis d'une étape finale d'élongation à 72°C pendant 5 min.

L'amplicon obtenu à la première amplification PCR a subi une seconde étape d'amplification PCR (PCR nichée) avec la paires d'amorce S/N/F4/+ /28054 et S/N/R4/- /29430 dans des conditions identiques à celles de la première amplification.

5 Le produit d'amplification obtenu correspondant à l'ADNc codant pour la protéine N de la souche de SARS-CoV issue du prélèvement n°031589a été séquencé à l'aide des amorces : S/N/F4/+ /28054, S/N/R4/- /29430, S/N/+ /28468, S/N/+ /28918 et S/N/- /28607 et cloné comme ci-dessus pour les autres ORF, pour
10 l'aide des amorces universelles M13 sens et M13 anti-sens, ainsi que des amorces S/N/+ /28468, S/N/+ /28918 et S/N/- /28607.

La séquence de l'amplicon représentant l'ADNc correspondant à l'ORF-N et incluant les ORF13 et ORF14 (SEQ ID NO :36) de la souche de SARS-CoV issue du prélèvement n°031589 ne comporte pas de différences par rapport aux
15 séquences correspondantes des isolats AY274119.3-Tor2 et AY278741-Urbani. La séquence de la protéine N de la souche de SARS-CoV issue du prélèvement n°031589 correspond à la séquence SEQ ID NO : 37 dans la liste de séquences jointe en annexe.

Les séquences des ORF13 et 14 de la souche de SARS-CoV issue du prélèvement n°031589 correspondent respectivement aux séquences SEQ ID NO : 32
20 et 34 dans la liste de séquences jointe en annexe.

Le plasmide dénommé SARS-N a été déposé sous le n° I-3048, le 5 juin 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient l'ADNc codant pour la protéine N de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle
25 que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 28054 à 29430 (SEQ ID NO :38), en référence à la séquence Genbank n° d'accès AY274119.3.

2.5) extrémités 5' et 3' non-codantes

a) extrémité 5' non-codante (5'NC)

30 a₁) synthèse de l'ADNc

Les ARN issus du prélèvement 031589, extraits comme ci-dessus, ont été soumis à une transcription inverse dans les conditions suivantes :

L'ARN (15 µl) et l'amorce S/L/-/443 (3 µl à la concentration de 5µM, ont été incubés 10 min à 75°C.

Ensuite, du Tampon de transcriptase inverse 5X (6 µl, INVITROGEN), des dNTP 10 mM (1 µl), du DTT 0,1M (3 µl) ont été ajoutés et le
5 mélange a été incubé à 50°C pendant 3 min.

Enfin la transcriptase inverse (3 µl de Superscript®, INVITROGEN) a été ajoutée au mélange précédent qui a été incubé à 50°C pendant 1h30 puis à 90 °C pendant 2 min.

L'ADNc ainsi obtenu a été purifié à l'aide du kit *QLAquick PCR purification* (QIAGEN), selon les recommandations du fabricant.

b₁) Réaction à la Terminal Transferase (TdT)

L'ADNc (10 µl) est incubé 2 min à 100°C, conservé dans la glace, puis sont ajoutés : H₂O (2,5 µl), tampon TdT 5X (4 µl, AMERSHAM), dATP 5mM (2 µl) et TdT (1,5 µl, AMERSHAM). Le mélange ainsi obtenu est incubé 45 min à
15 37°C puis 2 min à 65°C.

Le produit obtenu est amplifié par une première réaction PCR à l'aide des amorces : S/L/-/225-206 et ancre 14T : 5'-AGATGAATTCGGTACCTTTTTTTTTTTTTTTT-3' (SEQ ID NO :68). Les conditions de l'amplification sont les suivantes : une étape initiale de dénaturation à 94°C pendant 2 min est suivie de 10 cycles comprenant une étape de dénaturation à 94°C
20 pendant 10 sec, une étape d'hybridation à 45°C pendant 30 sec puis une étape d'élongation à 72°C pendant 30 sec puis de 30 cycles comprenant une étape de dénaturation à 94°C pendant 10 sec, une étape d'hybridation à 50°C pendant 30 sec puis une étape d'élongation à 72°C pendant 30 sec, puis d'une étape finale d'élongation à
25 72°C pendant 5 min.

Le produit de la première amplification PCR a subi une seconde étape d'amplification à l'aide des amorces : S/L/-/204-185 et ancre 14T précitée dans des conditions identiques à celles de la première amplification. L'amplicon ainsi obtenu a été purifié, séquencé à l'aide de l'amorce S/L/-/182-163 puis il a été cloné
30 comme ci-dessus pour les différentes ORF, pour donner le plasmide dénommé SARS-

5'NC. L'ADN de ce clone a été isolé et séquencé à l'aide des amorces universelles M13 sens et M13 anti-sens et de l'amorce S/L/-/182-163 précitée.

L'amplicon représentant l'ADNc correspondant à l'extrémité 5'NC de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589 correspond à la séquence SEQ ID NO : 72 dans la liste de séquences jointe en annexe ;
5 cette séquence ne comporte pas de différences par rapport aux séquences correspondantes des isolats AY274119.3-Tor2 et AY278741-Urbani.

Le plasmide dénommé SARS-5'NC a été déposé sous le n° I- 3124, le 7 novembre 2003, auprès de la Collection Nationale de Cultures de
10 Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15 ; il contient l'ADNc correspondant à l'extrémité 5'non codante du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant aux nucléotides des positions 1 à 204 (SEQ ID NO :39), en référence à la séquence Genbank n° d'accès AY274119.3.

15 b) extrémité 3'non-codante (3'NC)

a₁) synthèse de l'ADNc

Les ARN issus du prélèvement 031589, extraits comme ci-dessus, ont été soumis à une transcription inverse, selon le protocole suivant : le mélange réactionnel contenant : ARN (5 µl), H₂O (5 µl), tampon de transcriptase inverse 5X (4
20 µl), dNTP 5 mM (2 µl), Oligo 20T 5µM (2 µl), RNasin 40 U/ µl (0,5 µl) et RT-AMV 10 UI/ µl (1,5 µl, PROMEGA) a été incubé dans un thermocycleur, dans les conditions suivantes : 45 min à 42°C, 15 min à 55°C, 5 min à 95°C, puis il a été maintenu à +4°C.

L'ADNc obtenu a été amplifié par une première réaction PCR à
25 l'aide des amorces S/N/+ /28468-28487 et ancre 14T précitée. Les conditions de l'amplification sont les suivantes : une étape initiale de dénaturation à 94°C pendant 2 min est suivie de 10 cycles comprenant une étape de dénaturation à 94°C pendant 20 sec, une étape d'hybridation à 45°C pendant 30 sec puis une étape d'élongation à 72°C pendant 50 sec puis de 30 cycles comprenant une étape de dénaturation à 94°C
30 pendant 20 sec, une étape d'hybridation à 50°C pendant 30 sec puis une étape d'élongation à 72°C pendant 50 sec, puis d'une étape finale d'élongation à 72°C pendant 5 min.

Le produit de la première amplification PCR a subi une seconde étape d'amplification à l'aide des amorces S/N/+28933-28952 et ancre 14T précitée, dans des conditions identiques à celles de la première amplification. L'amplicon ainsi obtenu a été purifié, séquencé à l'aide de l'amorce S/N/+29257-29278 et cloné
5 comme ci-dessus pour les différentes ORF, pour donner le plasmide dénommé SARS-3'NC. L'ADN de ce clone a été isolé et séquencé à l'aide des amorces universelles M13 sens et M13 anti-sens et de l'amorce S/N/+29257-29278 précitée.

L'amplicon représentant l'ADNc correspondant à l'extrémité 3'NC de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589 correspond à la séquence SEQ ID NO :73 dans la liste de séquences jointe en annexe ; cette
10 séquence ne comporte pas de différences par rapport aux séquences correspondantes des isolats AY274119.3-Tor2 et AY278741-Urbani.

Le plasmide dénommé SARS-3'NC a été déposé sous le n° I-3123 le 7 novembre 2003, auprès de la Collection Nationale de Cultures de Microorganismes,
15 25 rue du Docteur Roux, 75724 Paris Cedex 15. ; il contient la séquence d'ADNc correspondant à l'extrémité 3'non codante du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, telle que définie ci-dessus, laquelle séquence correspondant à celle située entre le nucléotide en position 28933 à 29727 (SEQ ID NO :40), en référence à la séquence Genbank n° d'accès AY274119.3, se
20 termine par une série de nucléotides a.

2.6) ORF1a et ORF1b

L'amplification de la région 5' contenant les ORF1a et ORF1b du génome du SARS-CoV issu du prélèvement 031589 a été réalisée en pratiquant des réactions de RT-PCR suivies de PCR nichées selon les mêmes principes que ceux
25 précédemment décrits pour les autres ORF. Les fragments amplifiés sont chevauchants sur plusieurs dizaines de bases, permettant ainsi la reconstruction informatique de la séquence complète de cette partie du génome. En moyenne, les fragments amplifiés sont de deux kilobases.

14 fragments chevauchants dénommés L0 à L12 ont ainsi été
30 amplifiés à l'aide des amorces suivantes :

**Tableau II: Amorces utilisées pour l'amplification
de la région 5'(ORF1a et ORF1b)**

REGION AMPLIFIEE ET SEQUENCEE (ne tient pas compte des amorces)	Amorce sens RT-PCR	Amorce antisens RT-PCR	Amorce sens PCR nichée	Amorce antisens PCR nichée
L0 50-480	S/L0/F1/+30	S/L0/R1/-481		
L1 231-2240	S/L1/F1/+147	S/L1/R1/-2336	S/L1/F2/+211	S/L1/R2/-2241
L2 2156-4167	S/L2/F1/+2033	S/L2/R1/-4192	S/L2/F2/+2136	S/L2/R2/-4168
L3 3913-5324	S/L3bis/F1/+3850	S/L3bis/R1/-5365	S/L3bis/F2/+3892	S/L3bis/R2/-5325
L4b 4952-6023	S/L4b/F1/+4878	S/L4b/R1/-6061	S/L4b/F2/+4932	S/L4b/R2/-6024
L4 5325-7318	S/L4/F1/+5272	S/L4/R1/-7392	S/L4/F2/+5305	S/L4/R2/-7323
L5 7296-9156	S/L5/F1/+7111	S/L5/R1/-9253	S/L5/F2/+7275	S/L5/R2/-9157
L6 9053-11066	S/L6/F1/+8975	S/L6/R1/-11151	S/L6/F2/+9032	S/L6/R2/-11067
L7 10928-12962	S/L7/F1/+10883	S/L7/R1/-13050	S/L7/F2/+10928	S/L7/R2/-12963
L8 12835-14834	S/L8/F1/+12690	S/L8/R1/-14857	S/L8/F2/+12815	S/L8/R2/-14835
L9 14765-16624	S/L9/F1/+14688	S/L9/R1/-16678	S/L9/F2/+14745	S/L9/R2/-16625
L10 16534-18570	S/L10/F1/+16451	S/L10/R1/-18594	S/L10/F2/+16514	S/L10/R2/-18571
L11 18521-20582	S/L11/F1/+18441	S/L11/R1/-20612	S/L11/F2/+18500	S/L11/R2/-20583
L12 20338-22205.	S/L12/F1/+20279	S/L12/R1/-22229	S/L12/F2/+20319	S/L12/R2/-22206

Tous les fragments ont été amplifiés dans les conditions suivantes, excepté le fragment L0 qui a été amplifié comme décrit ci-dessus pour l'ORF-M :

- 5 - RT-PCR: 30 min à 42°C, 15 min à 55°C, 2 min à 94°C, puis l'ADNc obtenu est amplifié dans les conditions suivantes : 40 cycles comprenant : une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 58°C pendant 30 sec puis une étape d'élongation à 68°C pendant 1 min 30 sec, avec 5 sec d'élongation supplémentaire à chaque cycle, puis une étape finale d'élongation à 68°C pendant 7 min.

- 10 - PCR nichée : une étape initiale de dénaturation à 94°C pendant 2 min est suivie de 35 cycles comprenant : une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 60°C pendant 30 sec puis une étape d'élongation à 72°C pendant 1 min 30 sec, avec 5 sec d'élongation supplémentaire à chaque cycle, puis une étape finale d'élongation à 72°C pendant 7 min.
- 15

Les produits d'amplifications ont été séquencés à l'aide des amorces définies dans le Tableau III ci-après :

Tableau III : Amorces utilisées pour le séquençage de la région 5' (ORF1a et ORF1b)

5

Noms	Séquences (SEQ ID NO : 76 à 139)
S/L3/+14932	5'-CCACACACAGCTTGTGGATA-3'
S/L4/+16401	5'-CCGAAGTTGTAGGCAATGTC-3'
S/L4/+16964	5'-TTTGGTGCTCCTTCTTATTG-3'
S/L4/-16817	5'-CCGGCATCCAAACATAATTT-3'
S/L5/-17633	5'-TGGTCAGTAGGGTTGATTGG-3'
S/L5/-18127	5'-CATCCTTTGTGTCAACATCG-3'
S/L5/-18633	5'-GTCACGAGTGACACCATCCT-3'
S/L5/+17839	5'-ATGCGACGAGTCTGCTTCTA-3'
S/L5/+18785	5'-TTCATAGTGCCTGGCTTACC-3'
S/L5/+18255	5'-ATCTTGGCGCATGTATTGAC-3'
S/L6/-19422	5'-TGCATTAGCAGCAACAACAT-3'
S/L6/-19966	5'-TCTGCAGAACAGCAGAAGTG-3'
S/L6/-110542	5'-CCTGTGCAGTTTGTCTGTCA-3'
S/L6/+110677	5'-CCTTGTGGCAATGAAGTACA-3'
S/L6/+110106	5'-ATGTCATTTGCACAGCAGAA-3'
S/L6/+19571	5'-CTTCAATGGTTTGCCATGTT-3'
S/L7/-111271	5'-TGCGAGCTGTCATGAGAATA-3'
S/L7/-111801	5'-AACCGAGAGCAGTACCACAG-3'
S/L7/-112383	5'-TTTGGCTGCTGTAGTCAATG-3'
S/L7/+112640	5'-CTACGACAGATGTCCTGTGC-3'
S/L7/+112088	5'-GAGCAGGCTGTAGCTAATGG-3'
S/L7/+111551	5'-TTAGGCTATTGTTGCTGCTG-3'
S/L8/-13160	5'-CAGACAACATGAAGCACCAC-3'
S/L8/-113704	5'-CGCTGACGTGATATATGTGG-3'
S/L8/-14284	5'-TGCACAATGAAGGATACACC-3'
S/L8/+114453	5'-ACATAGCTCGCGTCTCAGTT-3'
S/L8/+113968	5'-GGCATTGTAGGCGTACTGAC-3'
S/L8/+113401	5'-GTTTGCAGGTGTAAGTGCAG-3'
S/L9/-15098	5'-TAGTGGCGGCTATTGACTTC-3'
S/L9/-15677	5'-CTAAACCTTGAGCCGCATAG-3'
S/L9/-16247	5'-CATGGTCATAGCAGCACTTG-3'
S/L9/+16323	5'-CCAGGTTGTGATGTCACTGAT-3'
S/L9/+15858	5'-CCTTACCCAGATCCATCAAG-3'
S/L9/+15288	5'-CGCAAACATAACACTTGCTG-3'
S/L10/-16914	5'-AGTGTTGGGTACAAGCCAGT-3'
S/L10/-17466	5'-GTTCCAAGGAACATGTCTGG-3'
S/L10/-18022	5'-AGGTGCCTGTGTAGGATGAA-3'
S/L10/+18245	5'-GGGCTGTCATGCAACTAGAG-3'
S/L10/+17663	5'-TCTTACACGCAATCCTGCTT-3'
S/L10/+17061	5'-TACCCATCTGCTCGCATAGT-3'
S/L11/-118877	5'-GCAAGCAGAATTAACCCTCA-3'
S/L11/-19396	5'-AGCACCACCTAAATTGCATC-3'
S/L11/-20002	5'-TGGTCCCTTTGAAGGTGTTA-3'
S/L11/+20245	5'-TCGAACACATCGTTTATGGA-3'
S/L11/+19611	5'-GAAGCACCTGTTTCCATCAT-3'
S/L11/+19021	5'-ACGATGCTCAGCCATGTAGT-3'
SARS/L1/F3/+800	5'-GAGGTGCAGTCACTCGCTAT-3'
SARS/L1/F4/+1391	5'-CAGAGATTGGACCTGAGCAT-3'

SARS/L1/F5/+1925	5'-CAGCAAACCACTCAATTCCT-3'
SARS/L1/R3/-1674	5'-AAATGATGGCAACCTCTTCA-3'
SARS/L1/R4/-1107	5'-CACGTGGTTGAATGACTTTG-3'
SARS/L1/R5/-520	5'-ATTTCTGCAACCAGCTCAAC-3'
SARS/L2/F3/+2664	5'-CGCATTGTCTCCTGGTTTAC-3'
SARS/L2/F4/+3232	5'-GAGATTGAGCCAGAACCAGA-3'
SARS/L2/F5/+3746	5'-ATGAGCAGGTTGTCATGGAT-3'
SARS/L2/R3/-3579	5'-CTGCCTTAAGAAGCTGGATG-3'
SARS/L2/R4/-2991	5'-TTTCTTCACCAGCATCATCA-3'
SARS/L2/R5/-2529	5'-CACCGTTCTTGAGAACAACC-3'
SARS/L3/F3/+4708	5'-TCTTTGGCTGGCTCTTACAG-3'
SARS/L3/F4/+5305	5'-GCTGGTGATGCTGCTAACTT-3'
SARS/L3/F5/+5822	5'-CCATCAAGCCTGTGTCGTAT-3'
SARS/L3/R3/-5610	5'-CAGGTGGTGCAGACATCATA-3'
SARS/L3/R4/-4988	5'-AACATCAGCACCATCCAAGT-3'
SARS/L3/R5/-4437	5'-ATCGGACACCATAGTCAACG-3'

Les séquences des fragments L0 à L12 de la souche de SARS-CoV issue du prélèvement répertorié sous le n° 031589, correspondent respectivement aux séquences SEQ ID NO :41 à SEQ ID NO :54 dans la liste de séquences jointe en annexe. Parmi ces séquences, seule celle correspondant aux fragments L5 comporte une différence nucléotidique par rapport à la séquence correspondante de l'isolat AY278741-Urbani. Cette mutation t/c en position 7919 aboutit à une modification de la séquence en acides aminés de la protéine correspondante, codée par l'ORF 1a: en position 2552, une valine (codon gtt ; AY278741) est changée en alanine (codon gct) dans la souche de SARS-CoV 031589. En revanche, aucune mutation n'a été identifiée par rapport à la séquence correspondante de l'isolat AY274119.3-Urbani. Les autres fragments ne présentent pas de différences par rapport aux séquences correspondantes des isolats Tor2 et Urbani.

Exemple 2 : Production et purification de protéines N et S recombinantes de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589

La protéine entière N et deux fragments polypeptidiques de la protéine S de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589 ont été produites chez *E. coli*, sous forme de protéines de fusion comprenant une étiquette polyhistidine N-ou C-terminale. Dans les deux polypeptides S, les séquences hydrophobes N et C-terminales de la protéine S (peptide signal : positions 1 à 13 et hélice transmembranaire : positions 1196 à 1218) ont été déléetées alors que l'hélice β (positions 565 à 687) et les deux motifs de type coiled-coils (positions 895 à 980 et 1155 à 1186) de la protéine S ont été préservés. Ces deux polypeptides sont

constitués par : un fragment long (S_L) correspondant aux positions 14 à 1193 de la séquence en acides aminés de la protéine S et un fragment court (S_C) correspondant aux positions 475 à 1193 de la séquence en acides aminés de la protéine S.

1) Clonage des ADNc N, S_L et S_C dans les vecteurs d'expression pIVEX2.3 et

5 pIVEX2.4

Les ADNc correspondant à la protéine N et aux fragments S_L et S_C ont été amplifiés par PCR dans des conditions standard, à l'aide de l'ADN polymérase Platinum Pfx® (INVITROGEN). Les plasmides SRAS-N et SRAS-S ont été utilisés comme matrice et les oligonucléotides suivants comme amorces :

- 10 5'-CCCATATGTCTGATAATGGACCCCAATCAAAC-3' (N sens, SEQ ID NO :55)
 5'-CCCCCGGGTGCCTGAGTTGAATCAGCAGAAGC-3' (N antisens, SEQ ID NO :56)
 5'-CCCATATGAGTGACCTTGACCGGTGCACCAC-3' (S_C sens, SEQ ID NO :57)
 5'-CCCATATGAAACCTTGACCCACCTGCTC-3' (S_L sens, SEQ ID NO :58)
 5'-CCCCCGGGTTTAATATATTGCTCATATTTTCCC-3' (S_C et S_L antisens, SEQ ID
 15 NO :59).

- Les amorces sens introduisent un site *NdeI* (souligné) alors que les amorces antisens introduisent un site *XmaI* ou *SmaI* (souligné). Les 3 produits d'amplification ont été purifiés sur colonne (kit *QIAquick PCR Purification*, QIAGEN) et clonés dans un vecteur approprié. L'ADN plasmidique purifié des 3 constructions
 20 (kit *QIAfilter Midi Plasmid*, QIAGEN) a été vérifié par séquençage et digéré par les enzymes *NdeI* et *XmaI*. Les 3 fragments correspondants aux ADNc N, S_L et S_C ont été purifiés sur gel d'agarose puis insérés dans les plasmides pIVEX2.3MCS (étiquette polyhistidine C-terminale) et pIVEX2.4d (étiquette polyhistidine N-terminale) préalablement digérés par les mêmes enzymes. Après vérification des constructions, les 6
 25 vecteurs d'expressions ainsi obtenus (pIV2.3N, pIV2.3 S_C , pIV2.3 S_L , pIV2.4N, pIV2.4 S_C également dénommé pIV2.4 S_L , pIV2.4 S_L) ont été ensuite utilisés, d'une part pour tester l'expression des protéines *in-vitro*, et d'autre part pour transformer la souche bactérienne BL21(DE3)pDIA17 (NOVAGEN). Ces constructions codent pour des protéines dont la masse moléculaire attendue est la suivante : pIV2.3N (47174
 30 Da), pIV2.3 S_C (82897 Da), pIV2.3 S_L (132056 Da), pIV2.4N (48996 Da), pIV2.4 S_L (81076 Da) et pIV2.4 S_L (133877 Da). Des bactéries transformées par pIV2.3N ont été déposées à la CNCM le 23 octobre 2003, sous le numéro I-3117, et des bactéries

transformées par pIV2.4S₁ ont été déposées à la CNCM le 23 octobre 2003, sous le numéro I-3118.

2) Analyse de l'expression des protéines recombinantes *in-vitro* et *in vivo*

L'expression de protéines recombinantes à partir des 6 vecteurs
5 recombinants a été testée, dans un premier temps, dans un système *in-vitro* (RTS100, Roche). Les protéines produites *in vitro*, après une incubation des vecteurs recombinants pIVEX, 4h à 30°C, dans le système RTS100, ont été analysées par western-blot à l'aide d'un anticorps anti-(his)₆ couplé à la peroxydase. Le résultat d'expression *in-vitro* (Figure 1) montre que seule la protéine N est exprimée en quantités importantes,
10 cela quelle que soit la position, N- ou C-terminale, de l'étiquette polyhistidine. Dans une seconde étape, l'expression des protéines N et S a été testée *in-vivo* à 30°C dans du milieu LB, en présence ou en l'absence d'inducteur (IPTG 1mM). La protéine N est très bien produite dans ce système bactérien (Figure 2) et se retrouve principalement dans une fraction soluble après lyse des bactéries. En revanche, la version longue
15 de S (S_L) est très peu produite et complètement insoluble (Figure 3). La version courte (S_C) présente également une très faible solubilité, mais un taux d'expression beaucoup plus élevé que celui de la version longue. Par ailleurs, la construction S_C fusionnée à une étiquette polyhistidine en position C-terminale présente une taille plus faible que celle attendue. Une expérience d'immunodétection avec un anticorps anti-poly-
20 histidine a montré que cette construction était incomplète. En conclusion, les deux constructions, pIV2.3N et pIV2.4S₁, exprimant respectivement la protéine N entière fusionnée à l'étiquette polyhistidine en C-terminal et la protéine S courte fusionnée à l'étiquette polyhistidine en N-terminal, ont été retenues pour produire les deux protéines en grande quantité afin de les purifier. Les plasmides pIV2.3N et pIV2.4S₁
25 ont été déposés respectivement sous le n° I-3117 et I-3118 auprès de la CNCM, 25 rue du Docteur Roux, 75724 PARIS 15, le 23 octobre 2003.

3) Analyse de l'activité antigénique des protéines recombinantes

L'activité antigénique des protéines N, S_L et S_C a été testée par western-blot, à l'aide de deux échantillons de sérum, provenant d'un même patient
30 infecté par le SARS-CoV, prélevés 8 jours (M12) et 29 jours-(M13) après le début des symptômes du SRAS. Le protocole expérimental est comme décrit à l'exemple 3. Les

résultats illustrés par la figure 4 montrent (i) la séroconversion du patient, et (ii) que la protéine N possède une plus forte réactivité antigénique que la protéine S courte.

4) Purification de la protéine N à partir de pIV2.3N

Plusieurs expériences de purification de la protéine N, produite à partir du vecteur pIV2.3N, ont été réalisées selon le protocole suivant. Les bactéries BL21(DE3)pDIA17, transformées par le vecteur d'expression pIV2.3N, ont été cultivées à 30°C dans 1 litre de milieu de culture contenant 0,1 mg/ml d'ampicilline, et induites par 1 mM IPTG quand la densité cellulaire, équivalente à $A_{600} = 0,8$, est atteinte (environ 3 heures). Après 2 heures de culture en présence d'inducteur, les cellules ont été récupérées par centrifugation (10 min à 5000 rpm), remises en suspension dans le tampon de lyse (50 mM NaH_2PO_4 , NaCl 0,3 M, 20 mM imidazole, pH 8 contenant le mélange d'inhibiteurs de protéases *Complete®*, Roche), et lysées par la presse de French (12000 psi). Après centrifugation du lysat bactérien (15 min à 12000 rpm), le surnageant (50 ml) a été déposé à un débit de 1ml/min sur une colonne (15 ml) de chélation métallique (Ni-NTA superflow, Qiagen), équilibrée par le tampon de lyse. Après lavage de la colonne par 200 ml de tampon de lyse, la protéine N a été éluée par un gradient d'imidazole (20 → 250 mM) en 10 volumes de colonne. Les fractions contenant la protéine N ont été rassemblées et analysées par électrophorèse en gel de polyacrylamide en conditions dénaturantes puis coloration au bleu de Coomassie. Les résultats illustrés par la figure 5 montrent que le protocole employé permet de purifier la protéine N avec une homogénéité très satisfaisante (95%) et un rendement moyen de 15 mg de protéine par litre de culture.

5) Purification de la protéine S_C à partir de pIV2.4S_C (pIV2.4S_I)

Le protocole suivi pour purifier la protéine S courte est très différent de celui décrit ci-dessus car la protéine est fortement agrégée dans le système bactérien (corps d'inclusion). Les bactéries BL21(DE3)pDIA17, transformées par le vecteur d'expression pIV2.4S_I ont été cultivées à 30°C dans 1 litre de milieu de culture contenant 0,1 mg/ml d'ampicilline, et induites par 1 mM IPTG quand la densité cellulaire, équivalente à $A_{600} = 0,8$, est atteinte (environ 3 heures). Après 2 heures de culture en présence d'inducteur, les cellules ont été récupérées par centrifugation (10 min à 5000 rpm), remises en suspension dans le tampon de lyse (0,1 M Tris-HCl, EDTA 1 mM, pH 7,5), et lysées par la presse de French (1200 psi). Après

centrifugation du lysat bactérien (15 min à 12000 rpm), le culot a été remis en suspension dans 25 ml de tampon de lyse contenant 2% Triton X100 et 10 mM β -mercaptoéthanol, puis centrifugé pendant 20 min à 12000 rpm. Le culot a été remis en suspension dans un tampon Tris-HCl 10 mM contenant 7 M urée, et mis en agitation douce pendant 30 min à température ambiante. Ce dernier lavage des corps d'inclusion avec 7 M urée est nécessaire pour éliminer la plupart des protéines membranaires d'*E. coli* qui co-sédimentent avec la protéine S_C agrégée. Après une dernière centrifugation pendant 20 min à 12000 rpm, le culot final est remis en suspension dans le tampon Tris-HCl 10 mM. L'analyse électrophorétique de cette préparation (Figure 6) montre que la protéine S courte peut être purifiée avec une homogénéité satisfaisante (environ 90%) à partir des corps d'inclusion (extrait insoluble).

Exemple 3 : Immunodominance de la protéine N

La réactivité des anticorps présents dans le sérum des patients atteints de pneumopathie atypique causée par le coronavirus associé au SRAS (SARS-CoV), vis-à-vis des différentes protéines de ce virus, a été analysée par western-blot dans les conditions décrites ci-après.

1) Matériel

a) lysat de cellules infectées par le SARS-CoV

Des cellules Vero E6 (2×10^6) ont été infectées par le SARS-CoV (isolat répertorié sous le numéro FFM/MA104) à une multiplicité d'infection (M.O.I.) de 10^{-1} ou 10^{-2} puis incubées dans du milieu DMEM contenant 2% de SVF, à 35°C dans une atmosphère contenant 5% de CO₂. 48 heures plus tard, le tapis cellulaire a été lavé avec du PBS puis lysé avec 500 μ l de tampon de dépôt préparé selon Laemmli et contenant du β -mercaptoéthanol. Les échantillons ont ensuite été bouillis 10 minutes puis soniqués 3 fois 20 secondes.

b) anticorps

b₁) sérum de patient atteint de pneumopathie atypique

Le sérum référencé au Centre National de Référence des virus influenzae (Région-Nord) sous le N° 20033168 est celui d'un patient français atteint d'une pneumopathie atypique causée par le SARS-CoV prélevé au jour 38 après le

début des symptômes ; le diagnostic d'infection par le SARS-CoV a été réalisé par RT-PCR nichée et PCR quantitative.

b₂) sérums polyclonaux de lapin monospécifiques dirigés contre la protéine N ou la protéine S

5 Les sérums sont ceux produits à partir des protéines recombinantes N et S_C (exemple 2), selon le protocole d'immunisation décrit à l'exemple 4 ; il s'agit du sérum du lapin P13097 (sérum anti-N) et du sérum du lapin P11135 (sérum anti-S).

2) Méthode

20 µl de lysat de cellules infectées par le SARS-CoV à des M.O.I. de
10 10⁻¹ et 10⁻² et, à titre de contrôle, 20 µl d'un lysat de cellules non infectées (mock) ont été séparés sur un gel SDS à 10% de polyacrylamide puis transférés sur une membrane de nitrocellulose. Après blocage dans une solution de PBS/lait 5%/Tween 0,1% et lavage en PBS/Tween 0,1%, cette membrane a été hybridée pendant une nuit à 4°C avec : (i) l'immun-sérum N° 20033168 dilué au 1/300, 1/1000 et 1/3000 dans le
15 tampon PBS/BSA 1%/Tween 0,1%, (ii) le sérum du lapin P13097 (sérum anti-N) dilué au 1/50000 dans le même tampon et (iii) le sérum du lapin P11135 (sérum anti-S) dilué au 1/10000 dans le même tampon. Après lavage en PBS/Tween, une hybridation secondaire a été réalisée à l'aide, soit d'anticorps polyclonaux de mouton dirigés contre les chaînes lourdes et légères des immunoglobulines G humaines et couplés à la
20 peroxidase (NA933V, Amersham), soit d'anticorps polyclonaux d'âne dirigés contre les chaînes lourdes et légères des immunoglobulines G de lapin et couplés à la peroxidase (NA934V, Amersham). Les anticorps fixés ont été révélés à l'aide du kit ECL+ (Amersham) et de films d'autoradiographie Hyperfilm MP (Amersham). Une échelle de masse moléculaire (kDa) est portée sur la figure.

25 3) Résultats

La figure 7 montre que trois polypeptides de masse moléculaire apparente 35, 55 et 200 kDa sont détectés spécifiquement dans les extraits de cellules infectées par le SARS-CoV.

Afin d'identifier ces polypeptides, deux autres immunoempreintes
30 (figure 8) ont été réalisées sur les mêmes échantillons et dans les mêmes conditions avec des anticorps polyclonaux de lapins spécifique de la nucléoprotéine N (lapin P13097, figure 8A) et de la protéine de spicule S (lapin P11135, figure 8B) Cette

expérience montre que le polypeptide de 200 kDa correspond à la glycoprotéine de spicule S du SARS-CoV, que le polypeptide de 55 kDa correspond à la nucléoprotéine N tandis que le polypeptide de 35 kDa représente vraisemblablement une forme tronquée ou dégradée de la N.

5 Les données présentées dans la figure 7 montrent donc que le sérum 20033168 réagit fortement avec la N et beaucoup plus faiblement avec la S du SARS-CoV, puisque les polypeptides de 35 et 55 kDa sont révélés sous la forme de bandes intenses pour des dilutions de 1/300, 1/1000 et 1/3000 de l'immunsérum alors que le polypeptide de 200 kDa n'est que faiblement révélé pour une dilution de 1/300. On
10 peut noter également qu'aucun autre polypeptide du SARS-CoV n'est détecté pour des dilutions supérieures au 1/300 du sérum 20033168.

Cette expérience indique que la réponse en anticorps spécifique de la N du SARS-CoV domine les réponses en anticorps spécifiques des autres polypeptides du SARS-CoV et en particulier la réponse en anticorps dirigée contre la
15 glycoprotéine S. Elle indique une immunodominance de la nucléoprotéine N lors des infections humaines par le SARS-CoV.

Exemple 4 : Préparation d'anticorps polyclonaux monospécifiques dirigés contre les protéines N et S du coronavirus associé au SRAS (SARS-CoV)

1) Matériel et méthode

20 Trois lapins (P13097, P13081, P13031) ont été immunisés avec le polypeptide recombinant purifié correspondant à l'intégralité de la nucléoprotéine (N), préparé selon le protocole décrit à l'exemple 2. Après une première injection de 0,35 mg par lapin de protéine émulsionnée en adjuvant complet de Freund (voie intra-dermique), les animaux ont reçus 3 injections de rappel à 3 puis 4 semaines
25 d'intervalle, de 0,35 mg de protéine recombinante émulsionnée en adjuvant incomplet de Freund.

Trois lapins (P11135, P13042, P14001) ont été immunisés avec le polypeptide recombinant correspondant au fragment court de la protéine S (S_C), produit comme décrit à l'exemple 2. Comme ce polypeptide est retrouvé principale-
30 ment sous la forme de corps d'inclusion dans le cytoplasme bactérien, les animaux ont reçus 4 injections intra-dermiques à 3-4 semaines d'intervalle d'une préparation de corps d'inclusion correspondant à 0,5 mg de protéine recombinante émulsionnée en

adjuvant incomplet de Freund. Les 3 premières injections ont été réalisées avec une préparation de corps d'inclusion préparés selon le protocole décrit à l'exemple 2, tandis que la quatrième injection a été réalisée avec une préparation de corps d'inclusion qui ont été préparés selon le protocole décrit à l'exemple 2 puis purifiés
5 sur gradient de saccharose et lavés en 2 % Triton X100.

Pour chaque lapin, un sérum pré-immun (p.i.) a été préparé avant la première immunisation et un immun-sérum (I.S.) 5 semaines après la quatrième immunisation.

Dans un premier temps, la réactivité des sérums a été analysée par
10 test ELISA vis à vis de préparations de protéines recombinantes semblables à celles utilisées pour les immunisations ; les tests ELISA ont été réalisés selon le protocole et avec les réactifs tels que décrits à l'exemple 6.

Dans un deuxième temps, la réactivité des sérums a été analysée en réalisant une immunoempreinte (western blot) d'un lysat de cellules infectées par le
15 SARS-CoV, en suivant le protocole tel que décrit à l'exemple 3.

2) Résultats

Les tests ELISA (figure 9) démontrent que les préparations de protéine N recombinante et de corps d'inclusion du fragment court de la protéine S (S_C) sont immunogènes chez l'animal et que le titre des sérums immuns est élevé (plus
20 de 1/25000).

L'immunoempreinte (figure 8) montre que le sérum immun du lapin P13097 reconnaît deux polypeptides présents dans les lysats de cellules infectées par le SARS-CoV : un polypeptide dont la masse moléculaire apparente (50-55 kDa selon les expériences) est compatible avec celle de la nucléoprotéine N (422 résidus, masse
25 moléculaire prédite de 46 kDa) et un polypeptide de 35 kDa, qui représente vraisemblablement une forme tronquée ou dégradée de la N.

Cette expérience montre également que le sérum du lapin P11135 reconnaît principalement un polypeptide dont la masse moléculaire apparente (180-220 kDa selon les expériences) est compatible avec une forme glycosylée de la S
30 (1255 résidus, chaîne polypeptidique non glycosylée de 139 kDa), ainsi que des polypeptides plus légers, qui représentent vraisemblablement des formes tronquées et/ou non glycosylées de la S.

En conclusion, l'ensemble de ces expériences démontrent que des polypeptides recombinants exprimés chez *E. coli* et correspondant aux protéines N et S du SARS-CoV permettent d'induire chez l'animal des anticorps polyclonaux capables de reconnaître les formes natives de ces protéines.

5 **Exemple 5 : Préparation d'anticorps polyclonaux monospécifiques dirigés contre les protéines M et E du coronavirus associé au SRAS (SARS-CoV)**

1) **Analyse de la structure des protéines M et E**

a) Protéine E

La structure de la protéine E du SARS-CoV (76 acides aminés) a été
10 analysée *in silico*, à l'aide de différents logiciels comme signalP v1.1, NetNGlyc 1.0, THMM 1.0 et 2.0 (Krogh et al., 2001, J. Mol. Biol., 305(3):567-580) ou encore TOPPRED (von Heijne, 1992, J. Mol. Biol. 225, 487-494). L'analyse montre que ce polypeptide non glycosylé est une protéine membranaire de type 1, contenant une
15 seule hélice transmembranaire (aa 12-34 d'après THMM), et dont la plus grande partie du domaine hydrophile (42 résidus) est localisée à l'extrémité C-terminale et vraisemblablement à l'intérieur de la particule virale (endodomaine). On peut noter une inversion dans la topologie prédite par les versions 1.0 (N-ter est externe) et 2.0 (N-ter est interne) du logiciel THMM, mais que d'autres algorithmes, notamment TOPPRED et THUMBUP (Zhou et Zhou, 2003, Protein Science 12 :1547-1555) confirment une
20 localisation externe de l'extrémité N-terminale de E.

b) Protéine M

Une analyse similaire réalisée sur la protéine M du SARS-CoV (221 acides aminés) montre que ce polypeptide ne possède pas de peptide signal (d'après le logiciel signalP v1.1) mais trois domaines transmembranaires (résidus 15-37, 50-72,
25 77-99 d'après THMM2.0) et un grand domaine hydrophile (aa 100-221) localisé à l'intérieur de la particule virale (endodomaine). Elle est vraisemblablement glycosylée sur l'asparagine en position 4 (d'après NetNGlyc 1.0).

Ainsi, en accord avec les données expérimentales connues pour les autres coronavirus, il est remarquable que les deux protéines M et E présentent des
30 endodomaines correspondant à la majeure partie des polypeptides et des ectodomaines de très petite taille.

- l'ectodomaine de E correspond vraisemblablement aux résidus 1 à 11 ou 1 à 12 de la protéine : MYSFVSEETGT(L), SEQ ID NO : 70. En effet, la probabilité associée à la localisation transmembranaire du résidu 12 est intermédiaire (0,56 d'après THMM 2.0).

5 - l'ectodomaine de M correspond vraisemblablement aux résidus 2 à 14 de la protéine : ADNGTITVEELKQ, SEQ ID NO : 69. En effet, la méthionine N-terminale de M est très probablement clivée du polypeptide mature car le résidu en position 2 est une Alanine (Varshavsky, 1996, 93:12142-12149).

Par ailleurs, l'analyse de l'hydrophobicité (Kyte & Doolittle, Hopp
10 & Woods) de la protéine E met en évidence que l'extrémité C-terminale de l'endodomaine de E est hydrophile et donc vraisemblablement exposée à la surface de ce domaine. Ainsi, un peptide synthétique correspondant à cette extrémité est un bon candidat immunogène pour induire chez l'animal des anticorps dirigés contre l'endodomaine de E. En conséquence, un peptide correspondant aux 24 résidus C-
15 terminaux de E a été synthétisé.

2) Préparation d'anticorps dirigés contre l'ectodomaine des protéines M et E et l'endodomaine de la protéine E

Les peptides M2-14 (ADNGTITVEELKQ, SEQ ID NO : 69), E1-
12 (MYSFVSEETGTL, SEQ ID NO : 70) et E53-76 (KPTVYVYSRV
20 KNLNSSEGVP DLLV, SEQ ID NO : 71) ont été synthétisés par Neosystem. Ils ont été couplés à la KLH (*Keyhole Limpet Hemocyanin*) à l'aide du MBS (m-maleimido-benzoyl-N-hydroxysuccinimide ester) via une cystéine ajoutée au cours de la synthèse soit en N-terminal du peptide (cas de E53-76) soit en C-terminal (cas de M2-14 et E1-
12).

25 Deux lapins ont été immunisés avec chacun des conjugués, en suivant le protocole d'immunisation suivant : après une première injection de 0,5 mg de peptide couplé à la KLH et émulsionné en adjuvant complet de Freund (voie intradermique), les animaux reçoivent 2 à 4 injections de rappel à 3 ou 4 semaines d'intervalle de 0,25 mg de peptide couplé à la KLH et émulsionné en adjuvant
30 incomplet de Freund.

Pour chaque lapin, un sérum pré-immun (p.i.) a été préparé avant la première immunisation et un immun-sérum (I.S.) est préparé 3 à 5 semaines après les injections de rappel.

La réactivité des sérums a été analysée par western blot à l'aide
5 d'extraits de cellules infectées par le SRAS-CoV (figure 43B) ou à l'aide d'extraits de cellules infectées par un virus recombinant de la vaccine exprimant la protéine E (VV-TG-E, figure 43A) ou M (VV-TN-M, figure 43C) de l'isolat 031589 du SRAS-CoV.

Les immunsérums des lapins 22234 et 22240, immunisés par le conjugué KLH-E53-76, reconnaissent un polypeptide d'environ 9 à 10kD, qui est
10 présent dans les extraits de cellules infectées par le SRAS-CoV mais absent dans les extraits de cellules non infectées (figure 43B). La masse apparente de ce polypeptide est compatible avec la masse prédite de la protéine E, qui est de 8,4 kD. De façon similaire, l'immunsérum du lapin 20047, immunisé par le conjugué KLH-E1-12, reconnaît un polypeptide présent dans les extraits de cellules infectées par le virus
15 VV-TG-E, dont la masse molaire apparente est compatible avec celle de la protéine E (figure 43A).

L'immunsérum des lapin 20013 et 20080, immunisés par le conjugué KLH-M2-14, reconnaît un polypeptide présent dans les extraits de cellules infectées par le virus VV-TN-M (figure 43C), dont la masse molaire apparente (18 kD
20 environ) est compatible avec celle de la glycoprotéine M, qui est de 25,1 kD et présente un point isoélectrique élevé (9.1 pour le polypeptide nu).

Ces résultats démontrent que les peptides E1-12 et E53-76 d'une part et le peptide M2-14 d'autre part permettent d'induire chez l'animal des anticorps polyclonaux capables de reconnaître les formes natives des protéines E et M
25 respectivement du SRAS-CoV.

Exemple 6 : Analyse de la réactivité en ELISA de la protéine N recombinante, vis-à-vis de sérums de patients atteints de SRAS

1) Matériel

L'antigène utilisé pour préparer les phases solides est la nucléo-
30 protéine N recombinante purifiée préparée selon le protocole décrit à l'exemple 2.

Les sérums à tester (Tableau IV) ont été choisis sur la base des résultats d'analyse de leur réactivité par immunofluorescence (titre IF-SRAS), vis-à-vis de cellules infectées par le SARS-CoV.

Tableau IV: Sérums testés en ELISA

Référence	N° sérum	Type de sérum	Date du Sérum***	Titre IF-SRAS
3050	A	Témoin	na*	nt**
3048	B	Témoin	na	nt
033168	D	Patient 1- SRAS	27/04/03 (J38)	320
033397	E	Patient-1 SRAS	11/05/03 (J52)	320
032632	F	Patient-2 SRAS	21/03/03 (J17)	2500
032791	G	Patient-3 SRAS	04/04/03 (J3)	<40
033258	H	Patient-3 SRAS	28/04/03 (J27)	160

*na : non-applicable. ** nt : non-testé. *** les dates indiquées correspondent au nombre de jours après le début des symptômes de SRAS.

2) Méthode

La protéine N (100 µl) diluée à différentes concentrations dans du tampon carbonate 0,1 M, pH 9,6 (1, 2 ou 4 µg/ml) est distribuée dans les puits de plaques ELISA, puis les plaques sont incubées une nuit à température du laboratoire. Les plaques sont lavées avec du tampon PBS-Tween, saturées avec du tampon PBS-lait écrémé-saccharose (5 %). Les sérums à tester (100 µl) préalablement dilués (1/50, 1/100, 1/200, 1/400, 1/800, 1/1600 et 1/3200) sont ajoutés, puis les plaques sont incubées 1 h à 37° C. Après 3 lavages, le conjugué anti-IgG humaines marqué à la peroxydase (référence 209-035-098, JACKSON) dilué au 1/18000 est ajouté puis les plaques sont incubées 1h à 37 °C. Après 4 lavages, le chromogène (TMB) et le substrat (H₂O₂) sont ajoutés et les plaques sont incubées 30min à température ambiante, à l'abri de la lumière. La réaction est ensuite arrêtée puis l'absorbance à 450 nm est mesurée à l'aide d'un lecteur automatique.

3) Résultats

Les tests ELISA (figure 10) démontrent que la préparation de protéine N recombinante est reconnue spécifiquement par les anticorps de sérums de patients atteints de SRAS prélevés en phase tardive de l'infection (≥ 17 jours après le début des symptômes) alors qu'elle n'est pas reconnue de façon significative par les anticorps d'un sérum de patient prélevé en phase précoce de l'infection (3 jours après le début des symptômes) ni par des sérums témoins de sujets non atteints de SRAS.

Exemple 7 : Tests ELISA élaborés pour une détection très spécifique et sensible d'une infection par le coronavirus associé au SRAS, à partir de sérums de patients

1) Test ELISA IgG indirect

a) Réactifs

Préparation des plaques

Les plaques sont sensibilisées par une solution de protéine N à 2µg/mL dans un tampon PBS 10mM pH 7,2, rouge de phénol à 0,25mL /L. 100 µL de solution sont déposés dans les puits et laissés incubés à température ambiante pendant une nuit. La saturation se fait par un prélavage en tampon PBS 10mM / tween 0,1%, suivi d'un lavage avec une solution de saturation PBS, 25% lait /saccharose.

Diluant serums

Tampon TRIS 0,48g/L, PBS 10mM, EDTA 3,7g/L, lait 15% v/v , pH 6,7

Diluant conjugué

Tampon citrate (15g/L), tween 0,5% , serum bovin 25%, NaCl 12%, lait écrémé 6% v/v PH 6,5

Conjugué

Conjugué anti-IgG humaines 50X, commercialisés par Bio-Rad : kit Platelia H. pylori ref 72778

Autres solutions :

Solution de lavage R2, Solutions de révélation au TMB R8 diluant, R9 chromogène, R10 solution d'arrêt : réactifs commercialisés commercialisées par Bio-Rad (ex : kit *Platelia pylori*, ref 72778)

b) Mode opératoire

Diluer les sérums au 1/200 dans le diluant échantillons
Distribuer 100µL/puits
Incubation 1h à 37°C
3 lavages en solution de LAVAGE R2 10x préalablement diluée 10 fois en eau déminéralisée (*i.e.*, solution de lavage 1X)

Distribuer 100µL de conjugué (conjugué 50x à diluer extemporanément dans le diluant conjugué fourni)

Incubation 1h à 37°C

4 lavages en solution de lavage 1X

- 5 distribuer 200µL/puits de solution de révélation (à diluer extemporanément ex: 1 mL de R9 dans 10mL de R8)

Incubation 30 min à température ambiante à l'obscurité

arrêter la réaction avec 100µL/puits de R10

LECTURE à 450/620nm

- 10 Les résultats peuvent être interprétés en prenant un serum SEUIL donnant une réponse au delà de laquelle les serums testés seront considérés comme positifs. Ce sérum est choisi et dilué de façon à donner un signal significativement supérieur au bruit de fond.

2) Test ELISA DOUBLE EPITOPE

- 15 a) Réactifs

Préparation des plaques

- 20 Les plaques sont sensibilisées par une solution de protéine N à 1µg/mL dans un tampon PBS 10mM pH 7,2, rouge de phénol à 0,25mL /L. 100 µL de solution sont déposés dans les puits et laissés incubés à température ambiante pendant une nuit. La saturation se fait par un prélavage en tampon PBS 10mM / 0,1% tween suivi d'un lavage avec une solution de saturation PBS 10mM, lait 25% (V/V)

- 25 *Diluant serums et conjugué*

Tampon TRIS saline 50mM pH8, lait 2%

Conjugué

Il s'agit de la protéine N recombinante purifiée, couplée à la peroxidase selon le protocole de Nakane (Nakane P.K. and Kawaoi A; (1974) :

- 30 *Peroxydase-labeled antibody, a new method of conjugation. The Journal of Histochemistry and Cytochemistry Vol22, N)23, pp 1084-1091.), dans des ratios molaires respectifs 1/2. Ce conjugué ProtN POD est utilisé à une concentration de 2µg/mL dans du diluant serum/conjugué.*

Autres solutions :

Solution de lavage R2, Solutions de révélation au TMB R8, diluant , R9 chromogène, R10 solution d'arrêt : réactifs commercialisées par Bio-Rad (ex kit platelia pylori ref 72778).

5

b) Mode opératoire*1ere étape en plaque de "prédilution"*

- Diluer chaque sérum au 1/5 dans la plaque de prédilution (48 µL de diluant +12
10 µL de sérum).
- Après avoir dilués l'ensemble des serums, distribuer 60µL de conjugué
- Le cas échéant, le mélange sérum + conjugué est laissé à incuber.

2eme étape en plaque de "réaction"

- 15 ▪ Transférer 100µL de mélange/puits dans la plaque de réaction
- Incubation 1h 37°C
- 5 lavages en solution de LAVAGE R2 10x préalablement diluée 10 fois en eau déminéralisée (--> solution de lavage 1x)
- distribuer 200µL/puits de solution de révélation (à diluer extemporanément ex:1
20 mL de R9 dans 10mL de R8)
- incubation 30min à température ambiante à l'abri de la lumière
- arrêter la réaction avec 100µL/puits de R10
- LECTURE à 450/620nm

25 De même que pour le test ELISA indirect, les résultats peuvent être interprétés en utilisant un serum "valeur seuil". Tout serum ayant une réponse supérieure au sérum valeur seuil sera considéré comme positif.

2) Résultats

Les sérums de patients classés comme cas probables de SRAS de l'hôpital français de Hanoi, Vietnam ou en relation avec l'hôpital français de Hanoi
30 (JYK) ont été analysés en utilisant le test IgG-N indirect et le test N double épitope.

Les résultats du test IgG-N indirect (figures 14 et 15) et N double épitope (figures 16 et 17) montrent une excellente corrélation entre eux ainsi qu'avec un test ELISA indirect comparant la réactivité des sérums vis-à-vis d'un lysat de

cellules VeroE6 non infectées ou infectées par le SRAS-CoV (ELISA-lysate SRAS-CoV ; voir le Tableau V ci-après). Tous les sérums prélevés 12 jours ou plus après le début des symptômes ont été trouvés positifs, y compris chez des patients pour lesquels l'infection par le virus du SRAS-CoV n'avait pas pu être documentée par analyse de prélèvements respiratoires par RT-PCR, vraisemblablement en raison d'un
 5 prélèvement trop tardif au cours de l'infection (\geq J12). Dans le cas du patient TTH pour lequel un prélèvement nasal réalisé à J7 a été trouvé négatif par RT-PCR, la qualité du prélèvement pourrait être en cause.

Certains sérums ont été trouvés négatifs alors que la présence de
 10 SRAS-CoV a été détectée par RT-PCR. Il s'agit dans tous les cas de sérums précoces prélevés moins de 10 jours après le début des symptômes (ex : sérum # 032637). Dans le cas d'un patient PTTH (sérum # 032673), seule une suspicion de SRAS était évoquée au moment où les prélèvements ont été réalisés.

En conclusion, les tests sérologiques IgG-N indirect et N-double
 15 épitope permettent de documenter l'infection par le SRAS-CoV chez tous les patients pour des sérums prélevés 12 jours ou plus après l'infection.

Tableau V : Résultats des tests ELISA

Num Pvt	Patient	Jour	PCR-SARS (1)	ELISA lysat SRAS-CoV (2)	IgG-N (2 ^{ème} série)	2Xépitope (2 ^{ème} série)
033168	JYK	38	POS	+++	>5000	NT
033597	JYK	74	POS	NT	\approx 5000	NT
032552	VTT	8	NEG-J3&J8&J12	NEG	<200	<5
032544	CTP	16	NEG J16&J20	++	>5000	>>20
032546	CJF	15	NEG J15&J19	++	>5000	>>20
032548	PTL	17	NEG J17&J21	++	>5000	>>20
032550	NTH	17	NEG-J17&J21	++	>5000	>>20
032553	VTT	8	NEG-J3&J8&J12	NEG	<200	<5
032554	NTBV	4	POS	NEG	<200	<5
032555	NTBV	4	POS	NEG	<200	
032564	NTP	15	POS	++	>5000	>>20
032629	NVH	4	POS	NEG	<200	<5
032631	BTTX	9	POS	NEG	<200	<5
032635	NHH	4	POS	NEG	<200	<5
032637	NHB	10	POS	NEG	<200	<5
032642	BTTX	9	POS	NEG	<200	<5

032643	LTDH	1	POS	NEG	<200	<5
032644	NTBV	4	POS	NEG	<200	<5
032646	TTH	12	NEG J7&J12&J16	++	>5000	>>20
032647	DTH	17	NEG J17&J21	++	>5000	>>20
032648	NNT	15	NEG J15&J19	++	>5000	>>20
032649	PTH	17	NEG J17&J21	++	>5000	>>20
032672	LVV	16	NEG J16&J20	+	>5000	>>20
032673	PTTH	NA	NEG	NEG	<200	<5
032674	PNB	17	NEG J17&J21	++	>5000	>>20
032682	VTH	12	NEG J12&J16	++	>5000	>>20
032683	DTV	17	NEG J17&J21	+	>1000	>>20

Notes:

(1) : Les analyses par RT-PCR ont été réalisées par RT-PCR nichée BNI, LC Artus et LC-N sur des écouvillonnages nasaux ou pharyngés; POS signifie qu'au moins un prélèvement a été trouvé positif chez ce patient.

- 5 (2) : La réactivité des sérums dans le test ELISA utilisant un lysat de cellules infectées par le SRAS-CoV a été classée en très fortement réactif (+++), fortement réactif (++), réactif (+) et négatif en fonction de la valeur DO obtenue aux dilutions testées

Exemple 8 : Détection du coronavirus associé au SRAS (SARS-CoV) par RT-PCR

- 10 1) Mise au point de conditions de RT-PCR en temps réel à l'aide d'amorces spécifiques du gène de la protéine de nucléocapside – test "Light Cycler N"

a) conception des amorces et des sondes

- La conception des amorces et sondes a été réalisée à partir de la séquence du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589, à l'aide du programme "Light Cycler Probe Design (Roche)".
- 15 Ainsi les deux séries d'amorces et de sondes suivantes ont été sélectionnées :

- série 1 (SEQ ID NO : 60, 61, 64, 65):

- amorce sens : N/+28507 : 5'-GGC ATC GTA TGG GTT G-3' [28507-28522]
- amorce antisens : N/-28774 : 5'-CAG TTT CAC CAC CTC C-3' [28774-28759]
- 20 - sonde 1 : 5'-GGC ACC CGC AAT CCT AAT AAC AAT GC-fluorescéine 3' [28561-28586]

- sonde 2 : 5' Red705 –GCC ACC GTG CTA CAA CTT CCT-phosphate [28588-28608]

- série 2 (SEQ ID NO : 62, 63, 66, 67)

- amorce sens : N/+ /28375 : 5'-GGC TAC TAC CGA AGA G-3' [28375-28390]

- amorce antisens : N/- /28702 : 5'-AAT TAC CGC GAC TAC G-3' [28702-28687]

5 - sonde 1 : SRAS/N/FL : 5'-ATA CAC CCA AAG ACC ACA TTG GC - fluorescéine 3' [28541-28563]

- sonde 2 : SRAS/N/LC705 : 5' Red705 -CCC GCA ATC CTA ATA ACA ATG CTG C-phosphate 3' [28565-28589]

10 b) analyse de l'efficacité des deux couples amorces

Afin de tester l'efficacité respective des deux couples d'amorces, une amplification par RT-PCR a été réalisée sur un ARN synthétique correspondant aux nucléotides 28054-29430 du génome de la souche de SARS-CoV issue du prélèvement répertorié sous le numéro 031589 et contenant la séquence du gène N.

15 De manière plus précise :

Cet ARN synthétique a été préparé par transcription *in vitro* à l'aide de l'ARN polymérase du phage T7, d'une matrice d'ADN obtenu par linéarisation du plasmide SRAS-N avec l'enzyme *Bam HI*. Après élimination de la matrice d'ADN par digestion à l'aide de DNase 1, les ARN synthétiques sont purifiés par une extraction
20 au phénol-chloroforme suivie de deux précipitations successives en acétate d'ammonium et isopropanol. Ils sont alors quantifiés par mesure de l'absorbance à 260 nm et leur qualité est contrôlée par le rapport des absorbances à 260 et 280 nm ainsi que par une électrophorèse en gel d'agarose. Ainsi, la concentration de la préparation d'ARN synthétique utilisée pour ces études est de 1,6 mg/ml, ce qui correspond à
25 $2,1.10^{15}$ copies/ml d'ARN.

Des quantités décroissantes d'ARN synthétique ont été amplifiées par RT-PCR à l'aide du kit "Superscript™ One-Step RT-PCR with Platinum® Taq" et les couples d'amorces n° 1 (N/+ /28507, N/- /28774) (figure 1A) et n° 2 (N/+ /28375, N/- /28702) (figure 1B), en suivant les indications du fournisseur. Les conditions d'ampli-
30 fication utilisées sont les suivantes : l'ADNc a été synthétisé par incubation 30 min à 45 °C, 15 min à 55°C puis 2 min à 94 °C puis il a été amplifié par 5 cycles comprenant : une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 45°C pendant 30 sec puis une étape d'élongation à 72°C pendant 30 sec, suivis de 35

cycles comprenant : une étape de dénaturation à 94°C pendant 15 sec, une étape d'hybridation à 55°C pendant 30 sec puis une étape d'élongation à 72°C pendant 30 sec, avec 2 sec d'élongation supplémentaire à chaque cycle, et d'une étape finale d'élongation à 72°C pendant 5 min. Les produits d'amplification obtenus ont ensuite
5 été maintenus à 10°C.

Les résultats présentés à la figure 11 montrent que le couple d'amorces n° 2 (N/+28375, N/-28702) permet de détecter jusqu'à 10 copies d'ARN (bande de faible intensité) ou 10² copies (bande de bonne intensité) contre 10⁴ copies pour le couple d'amorces n° 1 (N/+28507, N/-28774). Les amplicons sont respecti-
10 vement de 268 pb (couple 1) et de 328 pb (couple 2).

c) mise au point de la RT-PCR en temps réel

Une RT-PCR en temps réel a été mise au point à l'aide du couple d'amorces n°2 et du couple de sonde constitué par SRAS/N/FL et SRAS/N/LC705 (figure 2).

15 L'amplification a été réalisée sur un LightCycler™ (Roche) à l'aide du kit "Light Cycler RNA Amplification Kit Hybridization Probes " (référence 2 015 145, Roche) dans les conditions optimisées suivantes. Un Mélange réactionnel contenant : H₂O (6,8 µl), MgCl₂ 25 mM (0,8 µl, 4 µM final de Mg²⁺), mélange réactionnel 5X (4 µl), sonde SRAS/N/FL 3µM (0,5 µl, 0,075 µM final), sonde SRAS/N/LC705 3
20 µM (0,5 µl, 0,075 µM final), amorce N/+28375 10 µM (1 µl, 0,5 µM final), amorce N/-28702 10 µM (1 µl, 0,5 µM final), mélange d'enzyme (0,4 µl) et échantillon (ARN viral, 5 µl) a été amplifié en suivant le programme suivant :

- <u>Transcription inverse</u> :	50°C	10:00min	analysis mode: none
- <u>Dénaturation</u> :	95°C	30sec x1	analysis mode: none
25 - <u>Amplification</u> :	95°C	2sec	}
	50°C	15sec	analysis mode: <u>quantification*</u> } x45
	72°C	13sec	rampe thermique 2,0°C/sec }
- <u>refroidissement</u> :	40°C	30sec x1	analysis mode: none

*La mesure de fluorescence se fait à la fin de l'hybridation et à chaque cycle (en mode
30 SINGLE).

Les résultats présentés à la figure 12 montrent que cette RT-PCR en temps réel est très sensible puisqu'elle permet de détecter 10^2 copies d'ARN synthétique dans 100% des 5 échantillons analysés (29/29 échantillons dans 8 expériences) et jusqu'à 10 copies d'ARN dans 100% des 5 échantillons analysés (40/45 échantillons dans 8 expériences). Elle montre également que cette RT-PCR permet de détecter la présence du génome du SARS-CoV dans un échantillon et de quantifier le nombre de génomes présents. A titre d'exemple, l'ARN viral d'un stock de SARS-CoV cultivé sur cellules Vero E6 a été extrait à l'aide du kit "Qiaamp viral RNA extraction" (Qiagen), dilué à $0,05 \cdot 10^{-4}$ et analysé par RT-PCR en temps réel selon le protocole décrit ci-dessus; l'analyse présentée à la figure 12 montre que ce stock de virus contient $6,5 \cdot 10^9$ génomes –équivalents/ml (geq/ml), ce qui est tout à fait similaire à la valeur de $1,0 \cdot 10^{10}$ geq/ml mesurée à l'aide du kit "RealArt™ HPA-Coronavirus LC RT PCR Reagents" commercialisé par Artus.

2) Mise au point de conditions de RT-PCR nichée ciblant le gène de l'ARN polymérase – test "RT-PCR nichée CDC (Centers for Disease Control and Prevention) /IP"

a) Extraction de l'ARN viral

20 Echantillon clinique : QIAamp viral RNA Mini Kit (QIAGEN) selon les indications du fabricant, ou une technique équivalente. L'ARN est élué dans un volume de 60 µl.

b) RT-PCR nichée "SNE/SAR"

Première étape : RT-PCR couplée « SNE »

Le kit Invitrogen "Superscript™ One-Step RT-PCR with Platinum® Taq" a été utilisé, mais le kit "Titan" de Roche Boehringer peut lui être substitué avec des résultats similaires.

Oligonucléotides :

- 30 - SNE-S1 5' GGT TGG GAT TAT CCA AAA TGT GA 3'
- SNE-AS1 5'GCA TCA TCA GAA AGA ATC ATC ATG 3'

-> Taille attendue : 440 pb

1. Préparer un mix :

H₂O 6,5 µl

- | | | |
|---|---------------------|---------|
| | Reaction mix 2X | 12,5 µl |
| | Oligo SNE-S1 50 µM | 0,2 µl |
| | Oligo SNE-AS1 50 µM | 0,2 µl |
| | RNAsin 40 U/µl | 0,12 µl |
| 5 | RT/Platinum Taq mix | 0,5 µl |
2. A 20 µl du mix, ajouter 5 µl d'ARN et procéder à l'amplification sur thermocycleur (conditions ABI 9600) :
- 2.1 45°C 30 min.
55°C 15 min.
- 10 94°C 2 min.
- 2.2. 94°C 15 sec. }
45°C 30 sec. } x 5 cycles
72°C 30 sec. }
- 2.3. 94°C 15 sec. }
- 15 55°C 30 sec. } x 35 cycles
72°C 30 sec. + 2sec./cycle }
- 2.4. 72°C 5 min.
- 2.5 10°C ∞
- Conservation à +4°C.
- 20 La RNAsin (N2511/N2515) de Promega a été utilisée comme inhibiteurs de RNase.
- Des ARN synthétiques ont servi de témoin positif. A titre de contrôle, 10³, 10² et 10 copies d'ARN synthétique R_{SNE} ont été amplifiées dans chaque expérience.
- 25 *Seconde étape : PCR nichée "SAR"*
- Oligonucléotides :
- SAR1-S 5' CCT CTC TTG TTC TTG CTC GCA 3'
- SAR1-AS 5' TAT AGT GAG CCG CCA CAC ATG 3'
- > Taille attendue : **121 pb**
- 30 1. Préparer un mix :
- | | | |
|--|----------------|---------|
| | H2O | 35,8 µl |
| | Tampon Taq 10X | 5 µl |

	MgCl ₂ 25 mM	4 µl
	Mix dNTPs 5 mM	2 µl
	Oligo SAR1-S 50 µM	0,5 µl
	Oligo SAR1-AS 50 µM	0,5 µl
5	Taq ADN pol 5 U/µl	0,25 µl

L'*AmpliTaq DNA Pol* d'Applied Biosystems a été utilisée (tampon 10X sans MgCl₂, réf 27216601).

2. A 48 µl du mix, ajouter 2 µl du produit de la première PCR et procéder à l'amplification (conditions ABI 9600) :

- | | | | | |
|----|------|------|-----------------------|---------------|
| 10 | 2.1. | 94°C | 2 min. | |
| | 2.2. | 94°C | 30 sec. | } |
| | | 45°C | 45 sec. | } x 5 cycles |
| | | 72°C | 30 sec. | } |
| | 2.3. | 94°C | 30 sec. | } |
| 15 | | 55°C | 30 sec. | } x 35 cycles |
| | | 72°C | 30 sec. + 1sec./cycle | } |
| | 2.4. | 72°C | 5 min. | |
| | 2.5 | 10°C | ∞ | |

3. Analyser 10 µl du produit réactionnel sur gel "*low-melting*" (type Seakem GTG) à 3% d'agarose.

La sensibilité du test niché est en routine, dans les conditions décrites, de 10 copies d'ARN.

4. Les fragments peuvent ensuite être purifiés sur QIAquick PCR kit (QIAGEN) et séquencés avec les oligos SAR1-S et SAR1-AS.

25 3) Détection de l'ARN du SARS-CoV par PCR à partir de prélèvements respiratoires

a) Première étude comparative

Une étude comparative a été réalisée sur une série de prélèvements respiratoires reçus par le Centre National de Référence du Virus Influenzae (région nord) et susceptibles de contenir du SARS-CoV. Pour ce faire, l'ARN a été extrait des 30 prélèvements à l'aide du kit "Qiampr viral RNA extraction" (Qiagen) et analysé par

RT-PCR en temps réel, d'une part à l'aide des couples d'amorces et de sondes de la série n° 2 dans les conditions décrites ci-dessus d'une part, et d'autre part à l'aide du kit "LightCycler SARS-CoV quantification kit" commercialisé par Roche (référence 03 604 438). Les résultats sont résumés dans le Tableau VI ci-dessous. Ils montrent

5 que 18 des 26 prélèvements sont négatifs et 5 des 26 prélèvements sont positifs pour les deux kits, tandis qu'un prélèvement est positif pour le seul kit Roche et deux pour les seuls réactifs N "série2". En outre, pour 3 prélèvements (20032701, 20032712, 20032714) les quantités d'ARN détectés sont nettement supérieures avec les réactifs (sondes et amorces) de la série n°2. Ces résultats indiquent que les amorces et sondes

10 N "série2" sont plus sensibles pour la détection du génome du SARS-CoV dans des prélèvements biologiques que celles du kit actuellement disponible.

Tableau VI: Analyse par RT-PCR en temps réel des ARN extraits d'une série de prélèvements de 5 patients à l'aide des couples d'amorces et de sondes de la série n° 2 (N "série 2") ou du kit "LightCycler SARS-CoV quantification kit" (Roche).

15 Le type de prélèvement est indiqué ainsi que le nombre de copies de génome viral mesurées dans chacun des deux tests. NEG : RT-PCR négative.

Prélèvements n°	Patient	Type de prélèvement	KIT ROCHE	N "série2"
20033082	K	nasal	NEG	NEG
20033083	K	pharyngé	NEG	NEG
20033086	K	nasal	NEG	NEG
20033087	K	pharyngé	NEG	NEG
20032802	M	nasal	NEG	NEG
20032803	M	expectoration	NEG	NEG
20032806	M	nasal ou pharyngé	NEG	NEG
20031746ARN2	C	pharyngé	NEG	NEG
20032711	C	nasal ou pharyngé	39	NEG
20032910	B	nasal	NEG	NEG
20032911	B	pharyngé	NEG	NEG
20033356	V	expectoration	NEG	NEG
20033357	V	expectoration	NEG	NEG
20031725	K	asp. endotrachéale	NEG	150
20032657	K	asp. endotrachéale	NEG	NEG
20032698	K	asp. endotrachéale	NEG	NEG
20032720	K	asp. endotrachéale	3	5
20033074	K	selles	115	257
20032701	M	pharyngé	443	1676
20032702	M	expectoration	NEG	249
20031747ARN2	C	pharyngé	NEG	NEG
20032712	C	inconnu	634	6914
20032714	C	pharyngé	17	223
20032800	B	nasal	NEG	NEG
20033353	V	nasal	NEG	NEG
20033384	V	nasal	NEG	NEG

b) Deuxième étude comparative

Les performances de différentes méthodes de RT-PCR nichée et de RT-PCR en temps réel ont ensuite été comparées pour 121 prélèvements respiratoires de cas possibles de SRAS de l'hôpital français de Hanoi, Vietnam, réalisés entre le 4^{ème} et le 17^{ème} jour après le début des symptômes. Parmi ces prélèvements, 14 avaient été trouvés positifs lors d'un premier test utilisant la méthode de RT-PCR nichée ciblant l'ORF1b (codant pour la réplicase) telle que décrite initialement par le Bernhard Nocht Institute (RT-PCR nichée BNI). Des informations concernant ce test sont disponibles sur internet, à l'adresse http://www15.bni-hamburg.de/bni/bni2/neu2/getfile.acgi?area_engl=diagnostics&pid=4112.

Les différents tests comparés dans cette étude sont :

- la méthode de RT-PCR quantitative selon l'invention, avec les amorces et sonde N "série 2" décrite ci-dessus (colonne Light Cycler N),
- le test de RT-PCR nichée ciblant le gène de l'ARN polymérase décrit ci-dessus, développé par le CDC, le BNI et l'Institut Pasteur (RT-PCR nichée CDC/IP),
- le kit ARTUS de référence "HPA Corona LC RT-PCR Kit # 5601-02", qui est un test de RT-PCR en temps réel ciblant le gène ORF1b,
- le test de RT-PCR nichée du BNI, ciblant également le gène de l'ARN polymérase, mentionné ci-dessus.

Les inventeurs ont constaté :

- 1) une variabilité inter-test pour une même technique, liée à la dégradation de la préparation d'ARN lors de décongélations répétées, notamment pour les échantillons contenant les quantités d'ARN les plus faibles,
- 2) une sensibilité réduite de la RT-PCR nichée CDC/IP par rapport à la RT-PCR nichée BNI, et
- 3) une sensibilité comparable du test de RT-PCR quantitative selon l'invention (Light Cycler N) par rapport au test Light Cycler (LC) Artus.

Ces résultats, présentés dans le Tableau VII ci-dessous, montrent que le test par RT-PCR quantitative selon l'invention constitue un excellent complément –

ou une alternative – aux tests actuellement disponibles. En effet, le coronavirus lié au SRAS est un virus émergent, susceptible d'évoluer rapidement. En particulier, le gène de la RNA polymérase du coronavirus lié au SRAS, qui est ciblé dans la plupart des tests actuellement disponibles, peut recombiner avec celui d'autres coronavirus non
5 liés au SRAS. L'utilisation d'un test ciblant exclusivement ce gène pourrait alors conduire à l'obtention de faux négatifs.

Le test par RT-PCR quantitative selon l'invention ne cible pas la même région génomique que le kit ARTUS, puisqu'il cible le gène codant pour la protéine N. En réalisant un test de diagnostique ciblant deux gènes différents du
10 coronavirus lié au SRAS, on peut donc espérer s'affranchir de résultats de type faux négatifs, qui pourraient être dus à l'évolution génétique du virus.

De plus, il apparaît particulièrement avantageux de cibler le gène de la protéine de nucléocapside, car il est très stable, du fait de la forte pression de sélection liée aux contraintes structurales élevées concernant cette protéine.

Tableau VII: Comparaison de différentes méthodes d'analyse par amplification génique, à partir de 121 Prélèvements de cas probables de SRAS de l'hôpital français de Hanoi, Vietnam (épidémie 2003)

N° CNR	Type prélèvement (1)	Jour prélèvement	Patient	RT- PCR nichée CDC/IP	RT- PCR nichée BNI	kit Light Cycler Artus	Light Cycler N (IP)
107 prélèvements	N et P			Négatif	Négatif	Négatif	Négatif
032529	P	10	NHB	Négatif	Positif	Négatif	Négatif
032530	N	10	NHB	Positif	Positif	3,10E+01	4,20E+01
032531	P	7	LP	Positif	Positif	7,70E+00	3,10E+00
032534	N	15	BND	Positif	Positif	1,60E+00	Négatif
032600	P	4	NHH	Négatif	Positif	Négatif	1,30E+02
032612	P	17	NTS	Négatif	Positif	Négatif	Négatif
032688	P	9	BTX	Positif	Positif	Négatif	Négatif
032689	N	4	NVH	Positif	Positif	1,20E+01	2,30E+02
032690	P	4	NVH	Négatif	Positif	1,60E+00	Négatif
032727	P	8	NVH	Positif	Positif	2,30E+02	4,00E+02
032728	N	8	NVH	Positif	Positif	1,10E+03	1,60E+04
032729	P	14	NHB	Positif	Positif	5,90E+00	3,40E+01
032730	N	14	NHB	Positif	Positif	1,30E+02	4,80E+02
032741	P	8	NHH	Positif	Positif	2,10E+02	1,30E+02
positifs				10	14	10	9
fraction détectée des 14 positifs				71,4%	100,0%	71,4%	64,3%

- 5 (1) P= écouvillonnage pharyngé
N= écouvillonnage nasal

Exemple 9 : Obtention et caractérisation d'anticorps monoclonaux dirigés contre la protéine N

10 Des souris Balb C ont été immunisées à l'aide de la protéine N recombinante purifiée et leurs cellules spléniques fusionnées avec un myélome murin approprié selon les techniques de Köhler et Milstein.

Dix-neuf hybridomes sécréteurs d'anticorps anti N ont été présélectionnés et leurs immuno réactivités précisées. Ces anticorps reconnaissent
15 bien la protéine N recombinante (en ELISA) avec des intensités variables, ainsi que la protéine virale, naturelle N en ELISA et/ou en Western Blot. Les figures 18 à 20 montrent les résultats de ces tests pour 15 de ces 19 anticorps monoclonaux.

Les clones 12, 17, 28, 57, 72, 76, 86, 87, 98, 103, 146, 156, 166, 170, 199, 212, 218, 219 et 222, fortement réactifs, ont été sous clonés. Les études de
20 spécificité ont été poursuivies avec les outils appropriés afin de préciser les épitopes

reconnus et vérifier l'absence de réactivité vis-à-vis des autres Coronavirus humains et de certains virus respiratoires.

Les études de cartographie (*mapping*) épitopique (réalisées sur membrane spot, à l'aide de peptides chevauchants de 15 aa) et les études
5 supplémentaires réalisées sur la protéine N naturelle en Western Blot ont révélé l'existence de 4 groupes d'anticorps monoclonaux :

1° Anticorps monoclonaux spécifiques d'un épitope linéaire majeur en position N-ter (75-81, séquence : INTNSVP)

Le représentant de ce groupe est l'anticorps 156. L'hybridome
10 produisant cet anticorps a été déposé à la Collection Nationale de Cultures de Microorganismes (CNCM) de l'Institut Pasteur (Paris, France) le 1^{er} décembre 2004, sous le numéro I-3331. Ce même épitope est également reconnu par un sérum de lapin (polyclonal anti N) obtenu par immunisation classique à l'aide de cette même protéine N.

15 2° Anticorps monoclonaux spécifiques d'un épitope linéaire majeur situé en position centrale (position 217-224, séquence : ETALALL); les représentants de ce groupe sont les anticorps monoclonaux 87 et 166. L'hybridome produisant l'anticorps 87 a été déposé à la CNCM le 1^{er} décembre 2004, sous le numéro I-3328.

20 3° Anticorps monoclonaux spécifiques d'un épitope linéaire majeur situé en position C-terminale (position 403-408, séquence : DFFRQL), les représentants de ce groupe sont les anticorps 28, 57 et 143. L'hybridome produisant l'anticorps 57 a été déposé à la CNCM le 1^{er} décembre 2004, sous le numéro I-3330.

25 4° Anticorps monoclonaux spécifiques d'un épitope discontinu, conformationel. Ce groupe d'anticorps ne reconnaît aucun des peptides recouvrant la séquence de la protéine N, mais réagissent fortement sur la protéine naturelle non dénaturée. Le représentant de ce dernier groupe est l'anticorps 86. L'hybridome produisant cet anticorps a été déposé à la CNCM le 1^{er} décembre 2004, sous le numéro I-3329.

30 Le Tableau VIII ci-après résume les résultats de cartographie épitopique obtenus :

Tableau VIII : Mapping épitopique des anticorps monoclonaux

Anticorps	Epitope	Position	Région
28	DFSRQL Q	403 ... 408	C – Ter.

143	DFSRQL Q		
76	DFSRQL Q		
57	DFSRQL Q		
	FFGMS RI	315... 319	
146	LPQRQ	383 ...387	
166	ETALALLL	217 ... 224	centrale
87	ETALALL	217... 224	
156	INTNSGP	75... 81	N – Ter.
86	Conformationnel		
212	Conformationnel		
170	Conformationnel		

En outre, comme illustré notamment aux figures 18 et 19, ces anticorps ne présentent pas de réactivité en ELISA et/ou en WB, vis à vis de la protéine N du coronavirus humain 229 E.

5 **Exemple 10 : Combinaisons des anticorps monoclonaux pour le développement d'un test d'immunocapture sensible et spécifique de l'antigène viral N dans le sérum ou les fluides biologiques des patients contaminés par le virus SARS Co V**

Les anticorps listés ci-dessous ont été retenus en raison de leurs propriétés bien particulières pour une étude supplémentaire de capture et détection de la protéine virale N, dans le sérum des sujets ou patients.

Ces anticorps ont été produits en ascite sur souris, purifiés par chromatographie d'affinité et utilisés seuls ou en combinaison comme anticorps de capture, et comme anticorps signal.

Liste des anticorps sélectionnés pour le propos :

- 15 - Acm anti région C-ter (n° 28, 57, 143)
- Acm anti région centrale (n° 87, 166)
- Acm anti région N-ter (n° 156)
- Acm anti épitope discontinu, conformationnel (86)

1) Préparation des réactifs :

- 20 a) Plaques ELISA immunocapture

Les plaques sont sensibilisées avec les solutions d'anticorps à 5 µg/ml en tampon carbonate 0,1 M, pH.9,6. Les solutions (monovalentes ou plurivalentes) sont déposées sous volume de 100 µl dans les puits et incubées pendant une nuit à température ambiante. Ces plaques sont ensuite lavées en tampon PBS (10 mM pH 7,4 additionné de 0,1% de Tween 20) puis saturées avec une solution de PBS additionnée de 0,3% de BSA et de 5% de Saccharose). Les plaques sont ensuite séchées puis conditionnées dans un sachet en présence d'un dessicant. Elles sont prêtes à l'emploi.

b) Conjugués

Les anticorps purifiés ont été couplés à la peroxydase selon le protocole de Nakane (Nakane et al. - 1974, J.of Histo and cytochemistry, vol.22, p1084-1091) dans un rapport de une molécule d'IgG pour 3 molécules de peroxydase. Ces conjugués ont été purifiés par chromatographie d'exclusion et conservés concentrés (concentration comprise entre 1 à 2 mg/ml) en présence de 50% de glycerol et à -20°C. Ils sont dilués pour leur mise en œuvre dans les essais à la concentration finale de 1 ou 2 µg/ml en tampon PBS (pH 7,4) additionné de 1% de BSA.

c) Autres réactifs

- Sérums humains négatifs pour tous les marqueurs sériques des virus HIV, HBV, HCV et THLV
- Pool de sérums humains négatifs additionné de 0,5% de Triton X 100
- Ag viral inactivé : surnageant de culture virale inactivé par irradiation et inactivation vérifiée après mise en culture sur cellules sensibles - titre de la suspension avant inactivation environ 10^7 particules infectieuses par ml ou encore environ 5×10^9 particules virales physiques par ml d'antigène
- Les échantillons d'Ag dilués en sérum humain négatif : ces échantillons ont été préparés par dilution au 1:100^e puis par dilution en série de raison 5. Ces échantillons non infectieux miment des échantillons humains supposés contenir des concentrations faibles à très faibles de nucléoprotéine virale N. De tels échantillons ne sont pas accessibles pour les travaux de routine

- Solution de lavage R2, solution de révélation TMB R8, chromogène R9 et solution d'arrêt R10, sont les réactifs génériques commercialisés par Bio-Rad dans ses trousse ELISA (ex : trousse *Platelia Pylori* ref.72778).

5 2) Mode opératoire

Les échantillons de sérums humains surchargés en Ag viral inactivé sont distribués à raison de 100 µl par cupule, directement dans les plaques sensibilisées, prêtes à l'emploi puis incubés pendant 1 heure à 37°C (incubation Bio-
10 Rad IPS).

Le matériel non retenu par la phase solide est éliminé par 3 lavages (lavage avec solution R2 diluée, laveur automatique LP 35).

Les conjugués appropriés, dilués à la concentration finale de 1 ou 2 µg/ml sont distribués à raison de 100 µl par cupule et les plaques incubées à nouveau
15 pendant une heure à 37° C (incubation IPS).

L'excès de conjugué est éliminé par 4 lavages successifs (solution R2 diluée - laveur LP 35).

La présence de conjugué fixé sur les plaques est révélée après addition de 100 µl de solution de révélation préparée avant emploi (1 ml de R9 et 10
20 ml de R8) et après incubation pendant 30 minutes, à température ambiante et à l'abri de la lumière.

La réaction enzymatique est finalement bloquée par addition de 100 µl de réactif R10 (H2SO4 1N) dans toutes les cupules.

La lecture est effectuée à l'aide d'un lecteur de microplaque
25 approprié à double longueur d'onde (450/620 nm).

Les résultats peuvent être interprétés en utilisant comme valeur seuil provisoire la moyenne d'au moins deux contrôles négatifs multiplié par un facteur 2 ou encore la moyenne de 100 sérums négatifs additionnée d'un incrément correspondant à 6 SD (*Standart Deviation* calculée sur les 100 mesures individuelles).

30

3) Résultats

Différentes combinaisons anticorps de capture et anticorps signal ont été testées en se fondant sur les propriétés des anticorps sélectionnés, et en évitant les

combinaisons d'anticorps spécifiques des mêmes épitopes en phase solide et en conjugués.

Les meilleurs résultats ont été obtenus avec les 4 combinaisons listées ci-dessous. Ces résultats sont reproduits dans le tableau IX ci-après.

- 5 1. Combinaison F/28
Phase solide (Acm 166 + 87 région centrale) : conjugué anticorps 28 (C-ter)
2. Combinaison G/28
Phase solide (Acm 86 - épitope conformationnel) : conjugué anticorps 28 (C-ter)
3. Combinaison H/28
- 10 4. Phase solide (Acms 86, 166 et 87 région centrale et épitope conformationnel) :
 conjugué anticorps 28 (C-ter)
5. Combinaison H/28 + 87
Phase solide (Acms 86, 166 et 87 région centrale et épitope conformationnel) :
 conjugué mixte anticorps 28 (C-ter) et 87 (central)
- 15 6. Combinaison G/87
Phase solide (Acm 86 - épitope conformationnel) : conjugué anticorps 87 (région
centrale)

- Les 4 premières combinaisons présentent des performances équivalentes et reproduites, supérieures aux autres combinaisons mises en œuvre
- 20 (comme par exemple la combinaison G/87). Bien entendu, dans ces combinaisons, un anticorps monoclonal peut être remplacé par un autre anticorps reconnaissant le même épitope. Ainsi, on peut citer les variantes suivantes :

7. Variante de la combinaison F/28
Phase solide (Acm 87 uniquement) : conjugué anticorps 57 (C-ter)
- 25 8. Variante de la combinaison G/28
Phase solide (Acm 86 - épitope conformationnel) : conjugué anticorps 57 (C-ter)
9. Variante de la combinaison H/28
10. Variante de la combinaison H/28 + 87
Phase solide (Acms 86 et 87 région centrale et épitope conformationnel) : conjugué
mixte anticorps 57 (C-ter) et 87 (central)
- 30

Tableau IX : Contrôle d'immunoréactivité des Acm anti nucléoprotéines SARS CoV : densités optiques mesurées avec chaque combinaison d'anticorps, en fonction des dilutions de l'antigène viral inactivé.

N°	Dilution	F/28	G/28	G/87	H/28	H/28+87
0	1/100	5	5	3,495	3,900	5
1	1/500	3,795	3,814	1,379	3,702	3,804
2	1/2 500	2,815	2,950	0,275	3,268	2,680
3	1/12 500	0,987	1,038	0,135	1,374	0,865
4	1/62 500	0,404	0,348	0,125	0,480	0,328
5	1/312 500	0,285	0,211	0,123	0,240	0,215
6	Témoin	0,210	0,200	0,098	0,186	0,156
7	Témoin	0,269	0,153	0,104	0,193	0,202

5

La limite de détection de ces 4 essais expérimentaux correspond à la dilution d'antigène en sérum négatif 1:62500. Une extrapolation rapide laisse supposer la détection de moins de 10^3 particules infectieuses par ml de sérums.

De cette étude, il ressort que les anticorps les mieux appropriés pour la capture de la nucléoprotéine virale native sont les anticorps spécifiques de la région centrale et/ou d'un épitope conformationnel, l'un et l'autre étant des anticorps sélectionnés également pour leur forte affinité pour l'antigène natif.

Ayant déterminé les meilleurs anticorps pour la composition de la phase solide, les anticorps à retenir en priorité pour la détection des antigènes fixés sur la phase solide, sont les anticorps complémentaires spécifiques d'un épitope dominant en région C-ter. L'emploi de tout autre anticorps complémentaire, mais spécifique des épitopes localisés en région N-ter de la protéine conduit à des résultats moyens ou médiocres.

Exemple 11 : Systèmes d'expression eucaryotes de la protéine de spicule (S) du coronavirus associé au SRAS (SRAS-CoV)

1) Optimisation des conditions d'expression de la S du SRAS-CoV en cellules de mammifères.

Les conditions d'expression transitoire de la protéine de spicule (S) du SRAS-CoV ont été optimisées en cellules de mammifères (293T, VeroE6).

25

Pour cela, un fragment d'ADN contenant le cDNA de la S du SRAS-CoV a été amplifié par PCR à l'aide des oligonucléotides 5'-ATAGGATCCA CCATGTTTAT TTTCTTATTA TTTCTTACTC TCACT-3' et 5'-ATACTCGAGTT

ATGTGTAATG TAATTTGACA CCCTTG-3' à partir du plasmide pSARS-S (C.N.C.M. n° I-3059) puis inséré entre les sites BamH1 et Xho1 du plasmide pTRIPΔU3-CMV contenant un vecteur lentiviral TRIP (Sirven, 2001, Mol. Ther., 3, 438-448) pour obtenir le plasmide pTRIP-S. Le fragment BamH1 et Xho1 contenant
5 le cDNA de la S a ensuite été sous-cloné entre les BamH1 et Xho1 du plasmide d'expression eucaryote pcDNA3.1(+) (Clontech) pour obtenir le plasmide pcDNA-S. Le fragment Nhe1 et Xho1 contenant le cADN de la S a ensuite été sous-cloné entre les sites correspondants du plasmide d'expression pCI (Promega) pour obtenir le plasmide pCI-S. Les séquences WPRE du virus de l'hépatite de la marmotte
10 ("*Woodchuck Hepatitis Virus posttranscriptional regulatory element*") et les séquences CTE ("*constitutive transport element*") du rétrovirus simien de Mason-Pfizer ont été insérées dans chacun des deux plasmides pcDNA-S et pCI-S entre les sites Xho1 et Xba1 pour obtenir respectivement les plasmides pcDNA-S-CTE, pcDNA-S-WPRE, pCI-S-CTE et pCI-S-WPRE (figure 21). Le plasmide pCI-S-WPRE
15 a été déposé à la CNCM, le 22 novembre 2004, sous le numéro I-3323. Tous les inserts ont été séquencés à l'aide d'un kit BigDye Terminator v1.1 (Applied Biosystems) et d'un séquenceur automatique ABI377.

La capacité des plasmides construits à diriger l'expression de la S du SRAS-CoV dans des cellules de mammifères a été recherchée après transfection de
20 cellules VeroE6 (figure 22). Dans cette expérience, des monocouches de 5×10^5 cellules VeroE6 en boîtes de Pétri de 35mm ont été transfectées avec 2 µg des plasmides pcDNA (à titre de contrôle), pcDNA-S, pCI et pCI-S et 6 µl de réactif Eugene6 selon les indications du fabricant (Roche). Après 48 heures d'incubation à 37°C et sous 5% de CO₂, des extraits cellulaires ont été préparé en tampon de dépôt
25 selon Laemmli, séparés sur un gel SDS à 8% de polyacrylamide, puis transférés sur une membrane de PVDF (BioRad). La détection de cette immunoempreinte (« western blot ») a été réalisée à l'aide d'un sérum polyclonal de lapin anti-S (immun-sérum du lapin P11135 : cf exemple 4 ci-dessus) et d'anticorps polyclonaux d'âne dirigés contre les IgG de lapin et couplés à la peroxidase (NA934V, Amersham). Les anticorps fixés
30 ont été révélés par luminescence à l'aide du kit ECL+ (Amersham) et de films d'autoradiographie Hyperfilm MP (Amersham).

Cette expérience (figure 22) montre que le plasmide pcDNA-S ne permet pas de diriger l'expression de la S du SRAS-CoV à des niveaux détectables tandis que le plasmide pCI-S permet une expression faible, proche de la limite de détection, qui peut être mise en évidence lorsque le film est surexposé. Des résultats
5 similaires ont été obtenus lorsque l'expression de la S a été recherchée par immunofluorescence (données non montrées). Cette impossibilité de détecter une expression efficace de la S ne peut pas être imputée aux techniques de détection utilisées puisque la protéine S peut être mise en évidence à la taille attendue (180 kDa) dans un extrait de cellules infectées par le SRAS-CoV ou dans un extrait de cellules
10 VeroE6 infectées par le virus recombinant de la vaccine VV-TF7.3 et transfectées par le plasmide pcDNA-S. Dans cette dernière expérience, le virus VV-TF7.3 exprime l'ARN polymérase du phage T7 et permet la transcription cytoplasmique d'un ARN non coiffé susceptible d'être traduit efficacement. Cette expérience suggère que les défauts d'expression décrits ci-dessus sont dus à une incapacité intrinsèque du cADN
15 de la S à être exprimé efficacement lorsque l'étape de transcription en ARN messager est réalisée au niveau nucléaire.

Dans une seconde expérience, l'effet des signaux CTE et WPRE sur l'expression de la S a été recherchée après transfection de cellules VeroE6 (figure 23A) et 293T (figure 23B) et selon un protocole similaire à celui décrit ci-dessus.
20 Alors que l'expression de la S ne peut pas être mis en évidence après transfection des plasmides pcDNA-S-CTE et pcDNA-S-WPRE dérivés de pcDNA-S, l'insertion des signaux WPRE et CTE améliore fortement l'expression de la S dans le contexte du plasmide d'expression pCI-S.

Pour préciser ce résultat, une deuxième série d'expériences a été
25 réalisée, où l'immunoempreinte est révélée de façon quantitative par luminescence et acquisition sur un dispositif d'imagerie numérique (FluorS, BioRad). L'analyse des résultats obtenus avec le logiciel QuantityOne v4.2.3 (BioRad) montre que les séquences WPRE et CTE augmentent respectivement l'expression de la S d'un facteur 20 à 42 et 10 à 26 en cellules Vero E6 (Tableau X). En cellules 293T (Tableau X),
30 l'effet de la séquence CTE est plus modéré (4 à 5 fois) tandis que celui de la séquence WPRE reste important (13 à 28 fois).

Tableau X analyse quantitative de l'effet des signaux CTE et WPRE sur l'expression de la S du SRAS-CoV : Des extraits cellulaires ont été préparés 48 heures après transfection de cellules VeroE6 ou 293T par les plasmides pCI, pCI-S, pCI-S-CTE et pCI-S-WPRE et analysés par western blot comme décrit dans la légende de la figure 22. Le western blot est révélé par luminescence (ECL+, Amersham) et acquisition sur un dispositif d'imagerie numérique (FluorS, BioRad). Les niveaux d'expression sont indiqués en fonction d'une échelle arbitraire où la valeur de 1 représente le niveau mesuré après transfection du plasmide pCI-S.

Deux expériences indépendantes ont été réalisées pour chacun des deux types cellulaires. Dans l'expérience 1 sur cellules VeroE6, les transfections ont été réalisées en dupliqués et les résultats sont indiqués sous la forme de moyennes et écart-types des niveaux d'expression mesurés.

Plasmide	Cellule	exp. 1	exp. 2
PCI	VeroE6	0,0	0,0
pCI-S	VeroE6	1,0±0,1	1,0
pCI-S-CTE	VeroE6	9,8 ±0,9	26,4
pCI-S-WPRE	VeroE6	20,1±2,0	42,3
PCI	293T	0,0	0,0
PCI-S	293T	1,0	1,0
PCI-S-CTE	293T	4,6	4,0
PCI-S-WPRE	293T	27,6	12,8

En résumé, l'ensemble de ces résultats montre que l'expression dans

des cellules de mammifères du cADN de la S du SRAS-CoV sous la dépendance de séquences promotrices de l'ARN polymérase II, requiert, pour être efficace, la présence d'un signal d'épissage ainsi que de l'une ou l'autre des séquences WPRE et CTE.

2) Obtention de lignées stables permettant l'expression de la S du SRAS-CoV

Le cADN de la protéine S du SRAS-CoV a été cloné sous la forme d'un fragment BamH1-Xho1 dans le plasmide pTRIP Δ U3-CMV contenant un vecteur lentiviral défectif TRIP à DNA flap central (Sirven et al, 2001, Mol. Ther., 3 :438-448) pour obtenir le plasmide pTRIP-S (figure 24). La co-transfection transitoire selon Zennou et al (2000, Cell, 101 : 173-185) de ce plasmide, d'un plasmide d'encapsidation (p8.2) et d'un plasmide d'expression de la glycoprotéine d'enveloppe G du VSV (pHCMV-G) dans des cellules 293T a permis la préparation de pseudoparticules rétrovirales contenant le vecteur TRIP -S et pseudotypées par la protéine d'enveloppe G. Ces vecteurs TRIP-S pseudotypés ont été utilisés pour transduire des cellules 293T et FRhK-4 : aucune expression de la protéine S n'a pu être mise en évidence par western blot et immunofluorescence dans les cellules transduites (données non présentées).

Les cassettes d'expression optimales constituées du promoteur immédiat/précoce du virus CMV, d'un signal d'épissage, du cDNA de la S et de l'un ou l'autre des signaux post-transcriptionnels WPRE ou CTE décrites ci-dessus ont alors été substituées à la cassette EF1 α -EGFP du vecteur d'expression lentiviral défectif à DNA FLAP central TRIP Δ U3-EF1 α (Sirven et al, 2001, Mol. Ther., 3 :438-448) (figure 25). Ces substitutions ont été réalisées par une série de sous-clonages successifs des cassettes d'expression de S qui ont été excisées des plasmides pCT-S-CTE (BglII-Apa1) ou respectivement pCI-S-WPRE (BglII-Sal1) puis insérées entre les sites Mlu1 et Kpn1 ou respectivement Mlu1 et Xho1 du plasmide TRIP Δ U3-EF1 α pour obtenir les plasmides pTRIP-SD/SA-S-CTE et pTRIP-SD/SA-S-WPRE, déposés à la CNM, le 1^{er} décembre 2004, sous les numéros I-3336 et I-3334, respectivement. Des vecteurs pseudotypés ont été produits selon Zennou et al (2000, Cell, 101 : 173-185) et utilisés pour transduire des cellules 293T (10000 cellules) et FRhK-4 (15000 cellules) selon une série de 5 cycles successifs de transduction avec une quantité de vecteur correspondant à 25 ng (TRIP-SD/SA-S-CTE) ou 22 ng TRIP-SD/SA-S-WPRE) de p24 par cycle.

Les cellules transduites ont été clonées par dilution limite et une série de clones ont été analysés pour l'expression de la S du SRAS-CoV

qualitativement par immunofluorescence (données non montrées), puis quantitativement par western blot (figure 25) à l'aide d'un sérum polyclonal de lapin anti-S. Les résultats présentés dans la figure 25 montrent que les clones 2 et 15 de cellules FrhK4-s-CTE transduites par TRIP-SD/SA-S-CTE et les clones 4, 9 et 12 de
5 cellules FRhK4-S-WPRE transduites par TRIP-SD/SA-S-WPRE permettent l'expression de la S du SRAS-CoV à des niveaux respectivement faibles et modérés si on les compare à ceux que l'on peut observer au cours de l'infection par le SRAS-CoV.

En résumé, les vecteurs TRIP-SD/SA-S-CTE et TRIP-SD/SA-S-
10 WPRE permettent l'obtention de clones stables de cellules FRhK-4 et de façon similaire 293T exprimant la S du SRAS-CoV, alors que les essais réalisés avec le vecteur "de base" TRIP-S sont restés infructueux, ce qui démontre la nécessité d'un signal d'épissage ainsi que de l'une, ou l'autre, des séquences CTE et WPRE pour l'obtention de clones cellulaires stables exprimant la protéine S.

15 En outre, ces modifications du vecteur TRIP (insertion d'un signal d'épissage et d'un signal post-transcriptionnel comme CTE et WPRE) pourraient s'avérer intéressantes pour améliorer l'expression d'autres cDNA que celui de la S.

3) Obtention de lignées stables permettant l'expression d'une forme soluble de la S du SRAS-CoV. Purification de cet antigène recombinant

20 Un cDNA codant pour une forme soluble de la protéine S (Ssol) a été obtenu en fusionnant les séquences codant pour l'ectodomaine de la protéine (acides aminés 1 à 1193) à celles d'une étiquette (FLAG : DYKDDDDK) *via* un linker BspE1 codant pour le dipeptide SG. Pratiquement, pour obtenir le plasmide pcDNA-Ssol, un fragment d'ADN codant pour l'ectodomaine de la S du SRAS-CoV a été
25 amplifié par PCR à l'aide des oligonucléotides 5'-ATAGGATCCA CCATGTTTAT TTTCTTATTA TTTCTTACTC TCACT-3' et 5'-ACCTCCGGAT TTAATATATT GCTCATATTT TCCCAA-3' à partir du plasmide pcDNA-S, puis inséré entre les sites uniques BamH1 et BspE1 uniques d'un plasmide d'expression eucaryote pcDNA3.1(+) (Clontech) modifié contenant entre ses sites BamH1 et Xho1 la
30 séquence de l'étiquette FLAG :

```
// GGATCC ...nnn... TCC GGA GAT TAT AAA GAT GAC GAC GAT AAA TAA
    BamH1           S  G  D  Y  K  D  D  D  D  K  ter
CTCGAG //
Xho1
```

5

Les fragments Nhe1-Xho1 et BamH1-Xho1, contenant le cADN de la S, ont ensuite été excisés du plasmide pcDNA-Ssol, et sous-clonés entre les sites correspondants du plasmide pTRIP-SD/SA-S-CTE et du plasmide pTRIP-SD/SA-S-WPRE, respectivement, pour obtenir les plasmides pTRIP-SD/SA-Ssol-CTE et

10 pTRIP-SD/SA-Ssol-WPRE, déposés à la CNM, le 1^{er} décembre 2004, sous les numéros I-3337 et I-3335, respectivement.

Des vecteurs pseudotypés ont été produits selon Zennou et coll. (2000, Cell, 101 : 173-185) et utilisés pour transduire des cellules FRhK-4 (15000 cellules) selon une série de 5 cycles successifs de transduction (15000 cellules) avec

15 une quantité de vecteur correspondant à 24 ng (TRIP-SD/SA-Ssol-CTE) ou 40 ng TRIP-SD/SA-Ssol-WPRE) de p24 par cycle. Les cellules transduites ont été clonées par dilution limite et une série de 16 clones transduits par TRIP-SD/SA-Ssol-CTE et de 15 clones par TRIP-SD/SA-Ssol-WPRE ont été analysés pour l'expression du

20 polypeptide Ssol par western blot révélé par un anticorps monoclonal anti-FLAG (figure 26 et données non montrées), ainsi que par un ELISA-capture spécifique du polypeptide Ssol qui a été mis au point dans ce but (Tableau XI et données non montrées). Une partie du processus de sélection des meilleurs clones sécréteurs est montré dans la figure 26. L'ELISA-capture repose sur l'utilisation de phases solides

25 recouvertes d'anticorps polyclonaux de lapins immunisés par du SRAS-CoV purifié et inactivé. Ces phases solides permettent la capture du polypeptide Ssol sécrété dans les surnageants cellulaires, dont la présence est ensuite révélée par une série d'étapes impliquant successivement la fixation d'un anticorps monoclonal anti-FLAG (M2, SIGMA), d'anticorps biotinylés polyclonaux de lapin anti-IgG(H+L) de souris (Jackson) et d'un conjugué streptavidine-peroxidase (Amersham) puis l'addition de

30 chromogène et de substrat (TMB + H₂O₂, KPL).

- Tableau XI** : analyse de l'expression de polypeptide Ssol par des lignées cellulaires transduites par les vecteurs lentiviraux TRIP-SD/SA-Ssol-WPRE et TRIP-SD/SA-Ssol-CTE. La sécrétion du polypeptide Ssol a été recherchée dans le surnageant d'une série de clones cellulaires isolés après transduction de cellules FRhK-4 par les vecteurs lentiviraux TRIP-SD/SA-Ssol-WPRE et TRIP-SD/SA-Ssol-CTE. Les surnageants dilués au 1/50 ont été analysés par un test ELISA-capture spécifique de la S du SRAS-CoV.

Vecteur	Clone	DO (450 nm)
Témoin	-	0,031
TRIP-SD/SA-Ssol-CTE	CTE2	0,547
	CTE3	0,668
	CTE9	0,171
	CTE12	0,208
	CTE13	0,133
TRIP-SD/SA-Ssol-WPRE	WPRE1	0,061
	WPRE10	0,134

- La lignée cellulaire sécrétant les quantités les plus élevées de polypeptide Ssol dans le surnageant de culture est la lignée FRhK4-Ssol-CTE3. Elle a été soumise à une deuxième série de 5 cycles de transduction par le vecteur TRIP-SD/SA-Ssol-CTE dans des conditions similaires à celles décrites ci-dessus puis clonée. Le sous-clone sécrétant les quantités les plus élevés de Ssol a été sélectionné par une combinaison d'analyse par western blot et ELISA-capture : il s'agit du sous-clone FRhK4-Ssol-30, qui a été déposé à la CNCM, le 22 novembre 2004, sous le numéro I-3325.

- La lignée FRhK4-Ssol-30 permet la production et la purification en quantité du polypeptide recombinant Ssol. Dans une expérience type où les conditions expérimentales de croissance, de production et de purification ont été optimisées, les cellules de la lignée FRhK4-Ssol-30 sontensemencées en milieu de culture standard (DMEM sans pyruvate contenant 4.5g/l de glucose et supplémenté par 5% de SVF, 100 U/ml de pénicilline et 100 µg/ml de streptomycine) sous la forme d'une monocouche sous-confluente (1 million de cellules pour chaque 100 cm² dans 20 ml de milieu). A confluence, le milieu standard est remplacé par le milieu de sécrétion où

la quantité de SVF est abaissée à 0.5% et la quantité de milieu réduite à 16 ml pour chaque 100 cm². Le surnageant de culture est prélevé après 4 à 5 jours d'incubation à 35°C et sous 5% de CO₂. Le polypeptide recombinant Ssol est purifié à partir du surnageant par l'enchaînement d'étapes de filtration sur membrane de polyethersulfone (PES) de 0.1 µm, de concentration par ultrafiltration sur une membrane de PES de point de coupure 50kD, de chromatographie d'affinité sur matrice anti-FLAG avec élution par une solution de peptide FLAG (DYKDDDDK) à 100 µg/ml en TBS (Tris 50mM pH 7.4, 150 mMNaCl) puis de chromatographie de gel filtration en TBS sur billes de sephadex G-75 (Pharmacia). La concentration du polypeptide recombinant Ssol purifié a été déterminée par test micro-BCA (Pierce) puis ses caractéristiques biochimiques analysées.

L'analyse par gel SDS à 8 % d'acrylamide coloré au nitrate d'argent met en évidence un polypeptide majoritaire dont la masse moléculaire est d'environ 180kD et dont le degré de pureté peut être évalué à 98% (figure 27A). Par spectrométrie de masse SELDI-TOF (Cyphergeren), deux pics principaux sont mis en évidence : ils correspondent à des formes simplement, et doublement chargées d'un polypeptide majoritaire dont la masse moléculaire est ainsi déterminée à 182,6 ± 3,7 kD (figures 27B et C). Après transfert sur membrane Prosorb et rinçage en TFA 0,1%, l'extrémité N-terminale du polypeptide Ssol a été séquencée en phase liquide par dégradation d'Edman sur 5 résidus (ABI494, Applied Biosystems) et déterminée comme étant SDLDR (figure 27D). Ceci démontre que le peptide signal localisé à l'extrémité N-terminale de la protéine S du SRAS-CoV, composé des aa 1 à 13 (MFIFLLFLTLTSG) d'après une analyse réalisée avec le logiciel signalP v2.0 (Nielsen et al., 1997, *Protein Engineering*, 10 : 1-6), est clivé du polypeptide Ssol mature. Le polypeptide recombinant Ssol est donc constitué des acides aminés 14 à 1193 de la protéine S du SRAS-CoV fusionnés en C-terminal à une séquence SGDYKDDDDK contenant la séquence de l'étiquette FLAG (soulignée). L'écart entre la masse molaire théorique du polypeptide Ssol nu (132,0 kD) et la masse molaire réelle du polypeptide mature (182,6 kD) suggère que le polypeptide Ssol est glycosylé.

Une préparation de polypeptide Ssol purifié et dont la concentration protéique a été déterminée par test micro-BCA, permet de réaliser une gamme étalon

pour mesurer à l'aide du test ELISA-capture décrit plus haut les concentrations de Ssol présent dans les surnageants de culture et de revisiter les caractéristiques des lignées sécrétrices. Selon ce test, la lignée FRhK4-Ssol-CT3 sécrète 4 à 6 µg/ml de polypeptide Ssol tandis que la lignée FRhK4-Ssol-30 sécrète 9 à 13 µg/ml de Ssol après 4 à 5 jours de culture à confluence. En outre, le schéma de purification présenté plus haut permet en routine de purifier de 1 à 2 mg de polypeptide Ssol par litre de surnageant de culture.

Exemple 12 : Immunisation génique visant la protéine de spicule (S) du coronavirus associé au SRAS (SRAS-CoV)

10 L'effet d'un signal d'épissage et des signaux post-transcriptionnels WPRE et CTE a été analysé après immunisation génique de souris BALB/c (figure 28).

Pour cela, des souris BALB/c ont été immunisées à des intervalles de 4 semaines par injection dans le *tibialis anterior* d'une solution saline de 50 µg d'ADN plasmidique de pcDNA-S et pCI-S ainsi que, à titre de contrôle, par 50 µg d'ADN plasmidique de pcDNA-N (dirigeant l'expression de la N du SRAS-CoV) ou de pCI-HA (dirigeant l'expression de la HA du virus grippal A/PR/8/34) et les sérums immuns collectés 3 semaines après la 2ème injection. La présence d'anticorps dirigés contre la S du SRAS-CoV a été recherchée par ELISA indirect en utilisant comme 20 antigène un lysat de cellules VeroE6 infectées par le SRAS-CoV et à titre de contrôle, un lysat de cellules VeroE6 non infectées. Les titres (TI) en anticorps anti-SRAS-CoV sont calculés comme l'inverse de la dilution produisant une DO spécifique de 0,5 (différence entre DO mesurée sur lysat de cellules infectées et DO mesurée sur lysat de cellules non infectées) après révélation par un anticorps polyclonal anti-IgG de 25 souris couplé à la peroxidase (NA931V, Amersham) et du TMB additionné de H₂O₂ (KPL) (figure 28A).

Dans ces conditions, le plasmide d'expression pcDNA-S ne permet l'induction que de faibles titres d'anticorps dirigées contre la S du SRAS-CoV chez 3 souris sur 6 ($\text{LOG}_{10}(\text{TI})=1,9\pm0,6$) alors que le plasmide pcDNA-N permet l'induction d'anticorps anti-N à des titres élevés ($\text{LOG}_{10}(\text{TI})= 3,9\pm0,3$) chez tous les animaux, et les plasmides contrôles (pCI, pCI-HA) n'entraînent aucun anticorps détectable ($\text{LOG}_{10}(\text{TI})<1,7$). Le plasmide pCI-S muni d'un signal d'épissage permet l'induction

d'anticorps à des titres élevés ($\text{LOG}_{10}(\text{TI}) = 3,7 \pm 0,2$), qui sont environ 60 fois supérieurs à ceux observés après injection du plasmide pcDNA-S ($p < 10^{-5}$).

L'efficacité des signaux post-transcriptionnels a été étudiée en réalisant une étude dose-réponse des titres en anticorps anti-S induits chez la souris BALB/c en fonction de la quantité d'ADN plasmidique utilisé comme immunogène (2 µg, 10 µg et 50 µg). Cette étude (figure 28B) démontre que le signal post-transcriptionnel WPRE améliore fortement l'efficacité de l'immunisation génique lorsque de faibles doses d'ADN sont utilisées ($p < 10^{-5}$ pour une dose de 2 µg d'ADN et $p < 10^{-2}$ pour une dose de 10 µg), alors que l'effet du signal CTE reste marginal ($p = 0,34$ pour une dose de 2 µg d'ADN).

Enfin, les anticorps induits chez la souris après immunisation génique neutralisent l'infectivité du SRAS-CoV *in vitro* (figures 29A et 29B) à des titres qui sont en rapport avec les titres mesurés par ELISA.

En résumé, l'utilisation d'un signal d'épissage et du signal post-transcriptionnel WPRE du virus de l'hépatite de la marmotte améliore de façon considérable l'induction d'anticorps neutralisants dirigés contre le SRAS-CoV après immunisation génique à l'aide d'ADN plasmidique dirigeant l'expression du cADN de la S du SRAS-CoV.

Exemple 13 : Applications diagnostiques de la protéine S

La réactivité en ELISA du polypeptide recombinant Ssol a été analysée vis-à-vis de sérums de patients atteints de SRAS.

Les sérums de cas probables de SRAS testés ont été choisis sur la base des résultats (positifs ou négatifs) d'analyse de leur réactivité spécifique vis-à-vis des antigènes natifs du SRAS-CoV par test d'immunofluorescence sur cellules VeroE6 infectées par le SRAS-CoV et/ou par test ELISA indirect en utilisant comme antigène un lysat de cellules VeroE6 infectées par le SRAS-CoV. Les sérums de ces patients sont identifiés par un numéro d'ordre du Centre National de Référence des Virus Influenzae ainsi que par les initiales du patient et le nombre de jours écoulés depuis le début des symptômes. Tous les sérums de cas probables (cf Tableau XII) reconnaissent les antigènes natifs du SRAS-CoV, à l'exception du sérum 032552 du patient VTT, pour lequel l'infection par le SRAS-CoV n'a pas pu être confirmée par RT-PCR réalisée sur prélèvements respiratoires des jours 3, 8 et 12. Un panel de

sérums témoin a été utilisé à titre de contrôle (sérums TV): il s'agit de sérums prélevés en France avant l'épidémie de SRAS survenue en 2003.

Tableau XII : sérums de cas probables de SRAS

sérum	patient	jour de prélèvement
031724	JYK	7
033168	JYK	38
033597	JYK	74
032632	NTM	17
032634	THA	15
032541	PHV	10
032542	NIH	17
032552	VTT	8
032633	PTU	16
032791	JLB	3
033258	JLB	27
032703	JCM	8
033153	JCM	29

- 5 Des phases solides sensibilisées par le polypeptide recombinant Ssol ont été préparées par adsorption d'une solution de polypeptide Ssol purifié à 2 µg/ml en PBS dans les puits d'une plaque ELISA, puis les plaques sont incubées une nuit à 4°C et lavées avec du tampon PBS-Tween (PBS, 0,1% Tween20). Après saturation des plaques ELISA par une solution de PBS-lait écrémé à 10% (poids/volume) et
- 10 lavage en PBS-tween, les sérums à tester (100 µl) sont dilués au 1/400 dans du tampon PBS-lait écrémé-Tween (PBS, 3% lait écrémé, 0,1% Tween) puis ajoutés dans les puits de la plaque ELISA sensibilisée. Les plaques sont incubées 1h à 37°C. Après 3 lavages avec du tampon PBS-Tween, le conjugué anti-IgG humaines marqué à la
- 15 peroxydase (réf NA933V, Amersham) dilué au 1/4000 dans du tampon PBS-lait écrémé-Tween est ajouté, puis les plaques sont incubées une heure à 37°C. Après 6 lavages avec du tampon PBS-Tween, le chromogène (TMB) et le substrat (H₂O₂) sont ajoutés et les plaques sont incubées 10 minutes à l'abri de la lumière. La réaction est arrêtée par ajout d'une solution 1N de H₃PO₄, puis l'absorbance est mesurée à 450 nm avec un référence à 620nm.

Les tests ELISA (figure 30) démontrent que le polypeptide recombinant Ssol est reconnu spécifiquement par les anticorps sériques de patients atteints de SRAS prélevés en phase moyenne ou tardive de l'infection (≥ 10 jours après le début des symptômes alors qu'il n'est pas reconnu de façon significative par les anticorps sériques de deux patients (JLB et JCM) prélevés en phase précoce de l'infection (3 à 8 jours après le début des symptômes) ni par des sérums témoins de sujets non atteints de SRAS. Les anticorps sériques des patients JLB et JCM montrent une séroconversion entre les jours 3 et 27 pour le premier et 8 et 29 pour le second après le début des symptômes, ce qui confirme la spécificité de la réactivité de ces sérums vis-à-vis du polypeptide Ssol.

En conclusion, ces résultats démontrent que le polypeptide recombinant Ssol peut être utilisé comme antigène pour la mise au point d'un test ELISA de diagnostic sérologique de l'infection par le SRAS-CoV.

Exemple 14 : Applications vaccinales de la protéine S soluble recombinante

L'immunogénicité du polypeptide recombinant Ssol a été étudiée chez la souris.

Pour cela, un groupe de 6 souris a été immunisé à 3 semaines d'intervalle avec 10 μ g de polypeptide recombinant Ssol adjuvé par 1 mg d'hydroxyde d'aluminium (Alu-gel-S, Serva) dilué en PBS. Trois immunisations successives ont été réalisées et les sérums immuns ont été prélevés 3 semaines après chacune des immunisations (IS1, IS2, IS3). A titre de contrôle, un groupe de souris (groupe mock) a reçu de l'hydroxyde d'aluminium seul selon le même protocole.

Les sérums immuns ont été analysés par pool pour chacun des 2 groupes par ELISA indirect en utilisant un lysat de cellules VeroE6 infectées par le SRAS-CoV comme antigène et à titre de contrôle, un lysat de cellules VeroE6 non infectées. Les titres en anticorps anti-SRAS-CoV sont calculés comme l'inverse de la dilution produisant une DO spécifique de 0,5 après révélation par un anticorps polyclonal anti-IgG(H+L) de souris couplé à la peroxidase (NA931V, Amersham) et du TMB additionné de H₂O₂ (KPL). Cette analyse (figure 31) montre que l'immunisation par le polypeptide Ssol induit chez la souris dès la première immunisation des anticorps dirigés contre la forme native de la protéine de spicule du

SRAS-CoV présente dans le lysat de cellules VeroE6 infectées. Après 2 puis 3 immunisations, les titres en anticorps anti-S deviennent très élevés.

Les sérums immuns ont été analysés par pool pour chacun des 2 groupes pour leur capacité à séroneutraliser l'infectivité du SRAS-CoV. 4 points de séroneutralisation sur cellules FRhK-4 (100 TCID₅₀ de SRAS-CoV) sont réalisés pour chacune des dilutions de raison 2 testées à partir du 1/20. Le titre séroneutralisant est calculé selon la méthode de Reed et Munsch comme l'inverse de la dilution neutralisant l'infectivité de 2 cupules sur 4. Cette analyse montre que les anticorps induits chez la souris par le polypeptide Ssol sont neutralisants : les titres observés sont très élevés après 2 puis 3 immunisations (supérieurs à 2560 et 5120 respectivement, Tableau XIII).

Tableau XIII : induction d'anticorps dirigés contre le SRAS-CoV après immunisation avec le polypeptide recombinant Ssol. Les sérums immuns ont été analysés par pool pour chacun des 2 groupes pour leur capacité à séroneutraliser l'infectivité de 100 TCID₅₀ du SRAS-CoV sur cellules FRhK-4. 4 points sont réalisés pour chacune des dilutions de raison 2 testées à partir du 1/20. Le titre séroneutralisant est calculé selon la méthode de Reed et Munsch comme l'inverse de la dilution neutralisant l'infectivité de 2 cupules sur 4.

groupe	sérums	Ac. Neutralisants
Mock	pi	< 20
	IS1	< 20
	IS2	< 20
	IS3	< 20
Ssol	pi	< 20
	IS1	57
	IS2	> 2560
	IS3	> 5120

Les titres neutralisants observés chez les souris immunisées avec le polypeptide Ssol atteignent des niveaux très supérieurs aux titres observés par Yang et coll. chez la souris (2004, Nature, 428: 561-564) et à ceux observés par Buchholz chez le hamster (2004, PNAS 101 : 9804-9809), qui protègent respectivement la souris et le hamster de l'infection par le SRAS-CoV. Il est donc vraisemblable que les anticorps

neutralisants induits chez la souris après immunisation par le polypeptide Ssol protègent ces animaux contre l'infection par le SRAS-CoV.

Exemple 15 : Gène synthétique optimisé pour l'expression en cellules de mammifères de la protéine de spicule (S) du coronavirus associé au SRAS (SRAS-CoV).

1) Conception du gène synthétique

Un gène synthétique codant pour la protéine de spicule du SRAS-CoV a été conçu à partir du gène de l'isolat 031589 (plasmide pSARS-S, C.N.C.M. n° I-3059) de façon à permettre des niveaux d'expression élevés dans des cellules de mammifères et en particulier dans les cellules d'origine humaine.

Pour cela :

- l'usage des codons du gène sauvage de l'isolat 031589 a été modifié de façon à se rapprocher du biais observé chez l'homme et à améliorer l'efficacité de traduction du mRNA correspondant

- le contenu global en GC du gène a été augmenté de façon à prolonger la demi-vie du mRNA correspondant

- les motifs, éventuellement cryptiques, susceptibles d'interférer avec une expression efficace du gène ont été supprimés (sites donneurs et accepteurs d'épissage, signaux de polyadénylation, séquences très riches (>80%) ou très pauvres (<30%) en GC, séquences répétées, séquences impliquées dans la formation de structures secondaires de l'ARN, boîtes TATA)

- un deuxième codon STOP a été ajouté pour permettre une terminaison efficace de la traduction.

En outre, des motifs CpG ont été introduits dans le gène de façon à augmenter son immunogénicité comme vaccin à ADN. Afin de faciliter la manipulation du gène synthétique, deux sites de restriction BamH1 et Xho1 ont été placés de part et d'autre de la phase ouverte de lecture de la protéine S, et les sites de restriction BamH1, Xho1, Nhe1, Kpn1, BspE1 et Sal1 ont été évités dans le gène synthétique.

La séquence du gène synthétique conçu (gène 040530) est donnée en SEQ ID No: 140.

Un alignement du gène synthétique 040530 avec la séquence du gène sauvage de l'isolat 031589 du SRAS-CoV déposé à la C.N.C.M. sous le numéro I-3059 (SEQ ID NO : 4, plasmide pSRAS-S) est présenté dans la figure 32.

2) Constructions plasmidiques

5 Le gène synthétique SEQ ID No: 140 a été assemblé à partir d'oligonucléotides synthétiques et cloné entre les sites KpnI et SacI du plasmide pUC-Kana pour donner le plasmide 040530pUC-kana. La séquence nucléotidique de l'insert du plasmide 040530pUC-kana a été vérifiée par séquençage automatique (Applied).

10 Un fragment KpnI-XhoI contenant le gène synthétique 040530 a été excisé du plasmide 040530pUC-kana et sous-cloné entre les sites NheI et XhoI du plasmide d'expression pCI (Promega) pour obtenir le plasmide pCI-SSYNTH, déposé à la CNCM le 1^{er} décembre 2004, sous le numéro I-3333.

Un gène synthétique codant pour la forme soluble de la protéine S a
15 ensuite été obtenu en fusionnant les séquences synthétiques codant pour l'ectodomaine de la protéine S (acides aminés 1 à 1193) à celles de l'étiquette (FLAG : DYKDDDDK) via un linker BspEI codant pour le dipeptide SG. Pratiquement, un fragment d'ADN codant pour l'ectodomaine de la S du SRAS-CoV a été amplifié par PCR à l'aide des oligonucléotides 5'-ACTAGCTAGC GGATCCACCA
20 TGTTTCATCTT CCTG -3' et 5'- AGTATCCGGAC TTGATGTACT GCTCGTACTT GC-3' à partir du plasmide 040530pUC-kana, digéré par NheI et BspEI puis inséré entre les sites uniques NheI et BspEI du plasmide pCI-Ssol, pour donner le plasmide pCI-SCUBE, déposé à la CNCM le 1^{er} décembre 2004, sous le numéro I-3332. (Les plasmides pCI-Ssol, pCI-Ssol-CTE et pCI-Ssol-WPRE (déposé à la CNCM, le 22
25 novembre 2004, sous le numéro I-3324) avaient été précédemment obtenus par sous-clonage du fragment KpnI-XhoI excisé du plasmide pcDNA-Ssol (voir note technique de la DI 2004-106) entre les sites NheI et XhoI des plasmides pCI, pCI-S-CTE et pCI-S-WPRE respectivement.)

Les plasmides pCI-Scube et pCI-Ssol codent pour le même
30 polypeptide recombinant Ssol.

3) Résultats

La capacité du gène synthétique codant pour la protéine S à diriger efficacement l'expression de la S du SRAS-CoV dans des cellules de mammifères a été comparée à celle du gène sauvage après transfection transitoire de cellules de primates (VeroE6) et de cellules humaines (293T).

5 Dans l'expérience présentée par la figure 33 et au Tableau XIV, des monocouches de 5×10^5 cellules VeroE6 ou 7×10^5 cellules 293T en boîtes de Pétri de 35mm ont été transfectées avec 2 μ g des plasmides pCI (à titre de contrôle), pCI-S, pCI-S-CTE, pCI-S-WPRE et pCI-Ssynth et 6 μ l de réactif Fugene6 selon les indications du fabricant (Roche). Après 48 heures d'incubation à 37°C et sous 5% de
10 CO₂, des extraits cellulaires ont été préparés en tampon de dépôt selon Laemmli, séparés sur un gel SDS à 8% de polyacrylamide puis transférés sur une membrane de PVDF (BioRad). La détection de cette immunoempreinte (« western blot ») a été réalisée à l'aide d'un sérum polyclonal de lapin anti-S (immun-sérum du lapin P11135 : cf exemple 4 ci-dessus) et d'anticorps polyclonaux d'âne dirigés contre les
15 IgG de lapin et couplés à la peroxidase (NA934V, Amersham). L'immunoempreinte est révélée de façon quantitative par luminescence à l'aide du kit ECL+ (Amersham) et acquisition sur un dispositif d'imagerie numérique (FluorS, BioRad).

L'analyse des résultats obtenus avec le logiciel QuantityOne v4.2.3 (BioRad) montre que dans cette expérience, le plasmide pCI-Synth permet
20 l'expression transitoire de la protéine S à des niveaux élevés dans les cellules VeroE6 et 293T, alors que le plasmide pCI-S ne permet pas d'induire une expression à des niveaux suffisants pour être détectée. Les niveaux d'expression observés sont de l'ordre de 2 fois supérieurs à ceux observés avec le plasmide pCI-S-WPRE.

- Tableau XIV : utilisation d'un gène synthétique pour l'expression de la S du SRAS-CoV.** Des extraits cellulaires préparés 48 heures après transfection de cellules VeroE6 ou 293T par les plasmides pCI, pCI-S, pCI-S-CTE, pCI-S-WPRE et pCI-Ssynth ont été séparés sur un gel SDS à 8% d'acrylamide et analysés par western blot à l'aide d'un anticorps polyclonal de lapin anti-S et d'un anticorps polyclonal anti-IgG(H+L) de lapin couplé à la peroxidase (NA934V, Amersham). Le western blot est révélé par luminescence (ECL+, Amersham) et acquisition sur un dispositif d'imagerie numérique (FluorS, BioRad). Les niveaux d'expression de la protéine S ont été mesurés en quantifiant les 2 bandes majoritaires repérées sur l'image (voir figure 33) et sont indiqués en fonction d'une échelle arbitraire où la valeur de 1 représente le niveau mesuré après transfection du plasmide pCI-S-WPRE.

plasmide	VeroE6	293T
pCI	0,0	0,0
pCI-S	≤ 0,1	≤ 0,1
pCI-S-CTE	0,5	≤ 0,1
pCI-S-WPRE	1,0	1,0
pCI-Ssynth	1,8	1,9

- Dans un second temps, la capacité du gène synthétique Scube à diriger efficacement la synthèse et la sécrétion du polypeptide Ssol par des cellules de mammifères a été comparée à celle du gène sauvage après transfection transitoire de cellules de hamster (BHK-21) et de cellules humaines (293T).

- Dans l'expérience présentée par le Tableau XV, des monocouches de 6×10^5 cellules BHK-21 et de 7×10^5 cellules 293T en boîtes de Pétri de 35mm ont été transfectées avec 2 µg des plasmides pCI (à titre de contrôle), pCI-Ssol, pCI-Ssol-CTE, pCI-Ssol-WPRE et pCI-Scube et 6 µl de réactif Fugene6 selon les indications du fabricant (Roche). Après 48 heures d'incubation à 37°C et sous 5% de CO₂, les surnageants cellulaires ont été prélevés et analysés de façon quantitative pour la sécrétion du polypeptide Ssol par un test ELISA-capture spécifique du polypeptide Ssol.

- L'analyse des résultats montre que, dans cette expérience, le plasmide pCI-Scube permet l'expression du polypeptide Ssol à des niveaux 8 fois (cellules BHK-21) à 20 fois (cellules 293T) plus élevés que le plasmide pCI-Ssol. Les niveaux d'expression observés sont de l'ordre de 2 fois (cellules 293T) à 5 fois (cellules BHK-21) supérieurs à ceux observés avec le plasmide pCI-Ssol-WPRE.

Tableau XV : utilisation d'un gène synthétique pour l'expression du polypeptide Ssol. Les surnageants ont été récoltés 48 heures après transfection de cellules BHK ou 293T par les plasmides pCI, pCI-Ssol, pCI-Ssol-CTE, pCI-Ssol-WPRE et pCI-Scube et analysés de façon quantitative pour la sécrétion du polypeptide Ssol par un test ELISA-capture spécifique du polypeptide Ssol. Les transfections ont été réalisées en dupliqués et les résultats sont indiqués sous la forme de moyennes et écart-types des concentrations de polypeptide Ssol (ng/ml) mesurées dans les surnageants.

Plasmide	BHK	293T
Pci	< 20	< 20
pCI-Ssol	< 20	56 ± 10
pCI-Ssol-CTE	< 20	63 ± 8
pCI-Ssol-WPRE	28 ± 1	531 ± 15
PCI-Scube	152 ± 6	1140 ± 20

En résumé, ces résultats montrent que l'expression dans des cellules de mammifères du gène synthétique 040530 codant pour la S du SRAS-CoV sous la dépendance de séquences promotrices de l'ARN polymérase II est bien plus efficace que celle du gène sauvage de l'isolat 031589. Cette expression est même plus efficace que celle dirigée par le gène sauvage en présence des séquences WPRE du virus de l'hépatite de la marmotte.

4) Applications

L'utilisation du gène synthétique 040530 codant pour la S du SRAS-CoV ou de sa variante Scube codant pour le polypeptide Ssol est susceptible de remplacer avantageusement le gène sauvage dans de nombreuses applications où l'expression de la S est nécessaire à des niveaux élevés. En particulier pour :

- améliorer l'efficacité de l'immunisation génique par des plasmides du type pCI-Ssynth voire pCI-Ssynth-CTE ou pCI-Ssynth-WPRE
- établir de nouvelles lignées cellulaires exprimant des quantités plus élevées de la protéine S ou du polypeptide Ssol à l'aide de vecteurs lentiviraux recombinants porteurs du gène Ssynth ou du gène Scube respectivement
- améliorer l'immunogénicité des vecteurs lentiviraux recombinants permettant l'expression de la protéine S ou du polypeptide Ssol
- améliorer l'immunogénicité de vecteurs vivants permettant l'expression de la protéine S ou du polypeptide Ssol comme des virus recombinants de la vaccine ou des virus rougeole recombinants (voir exemples 16 et 17 ci-après)

Exemple 16 : Expression de la protéine de spicule (S) du coronavirus associé au SRAS (SRAS-CoV) à l'aide de virus recombinants de la Vaccine.

Application vaccinale.

Application à la production d'une forme soluble de la protéine de spicule (S) et conception d'un test de sérologie du SRAS.

1) Introduction

Le but de cet exemple est d'évaluer la capacité de virus recombinants de la vaccine (VV) exprimant différents antigènes du coronavirus associé au SRAS (SRAS-CoV) à constituer de nouveaux candidats vaccins contre le SRAS et un moyen de produire des antigènes recombinants en cellules de mammifères.

Pour cela, les inventeurs se sont intéressés à la protéine de spicule (S) du SRAS-CoV, qui permet d'induire après immunisation génique chez l'animal des anticorps neutralisants l'infectivité du SRAS-CoV, ainsi qu'à une forme soluble et sécrétée de cette protéine, le polypeptide Ssol, qui est composé de l'ectodomaine (aa 1-1193) de la S fusionné en son extrémité C-ter à une étiquette FLAG (DYKDDDDK) via un linker BspE1 codant pour le dipeptide SG. Ce polypeptide Ssol présente une antigénicité similaire à celle de la protéine S et permet, après injection à la souris sous la forme d'une protéine purifiée adjuvée en hydroxide d'aluminium, l'induction de titres élevés d'anticorps neutralisants contre le SRAS-CoV.

Les différentes formes du gène S ont été placées sous la dépendance du promoteur du gène 7.5K puis introduites au sein du locus de la thymidine kinase (TK) de la souche Copenhague du virus de la vaccine par double recombinaison homologue *in vivo*. Afin d'améliorer l'immunogénicité des virus vaccine recombinants, un promoteur tardif synthétique a été choisi à la place du promoteur 7.5K, pour augmenter la production de S et Ssol au cours des phases tardives du cycle viral.

Après avoir isolé les virus vaccine recombinants et vérifié leur capacité à exprimer l'antigène S du SRAS-CoV, leur capacité à induire chez la souris une réponse immunitaire contre le SRAS a été testée. Après avoir purifié l'antigène Ssol du surnageant de cellules infectées, un test ELISA de sérodiagnostic du SRAS a été conçu, et son efficacité a été évaluée à l'aide de sérums de cas probables de SRAS.

2) Construction des virus recombinants

Des virus recombinants de la vaccine dirigeant l'expression de la glycoprotéine S de l'isolat 031589 du SRAS-CoV et d'une forme soluble et sécrétée de cette protéine, le polypeptide Ssol, sous la dépendance du promoteur 7.5K ont été obtenus. Dans le but d'augmenter les niveaux d'expression de S et Ssol, des virus recombinants dans lesquels les cDNA de S et de Ssol sont placés sous la dépendance d'un promoteur synthétique tardif ont également été obtenus.

Le plasmide pTG186poly est un plasmide de transfert pour la construction de virus recombinants de la vaccine (Kieny, 1986, Biotechnology, 4 :790-795). A ce titre, il contient le gène de la thymidine kinase du VV dans lequel a été inséré le promoteur du gène 7.5K suivi d'un site multiple de clonage permettant l'insertion de gènes hétérologues (figure 34A). Le promoteur du gène 7.5K contient en fait un tandem de deux séquences promotrices actives respectivement durant les phases précoces (P_E) et tardives (P_L) du cycle de réplication du virus de la vaccine. Les fragments BamH1-Xho1 ont été excisés des plasmides pTRIP-S et pcDNA-Ssol respectivement et insérés entre les sites BamH1 et Sma1 du plasmide pTG186poly pour donner les plasmides pTG-S et pTG-Ssol (figure 34A). Les plasmides pTG-S et pTG-Ssol ont été déposés à la CNCM, le 2 décembre 2004, sous les numéros I-3338 et I-3339, respectivement.

Les plasmides pTN480, pTN-S et pTN-Ssol ont été obtenus à partir des plasmides pTG186poly, pTG-S et pTG-Ssol respectivement, en substituant le fragment Nde1-Pst1 contenant le promoteur 7.5K par un fragment d'ADN contenant le promoteur tardif synthétique 480, qui a été obtenu par hybridation des oligonucléotides 5'- TATGAGCTTT TTTTTTTTTT TTTTTTTGGC ATATAAATAG ACTCGGCGCG CCATCTGCA-3' et 5'- GATGGCGCGC CGAGTCTATT TATATGCCAA AAAAAAAAAA AAAAAAAGC TCA-3' (figure 34B). L'insert a été séquencé à l'aide d'un kit BigDye Terminator v1.1 (Applied Biosystems) et d'un séquenceur automatique ABI377. La séquence du promoteur synthétique tardif 480 tel que cloné dans les plasmides de transfert de la série pTN est indiquée figure 34C. Les plasmides pTN-S et pTN-Ssol ont été déposés à la CNCM, le 2 décembre 2004, sous les numéros I-3340 et I-3341, respectivement.

Les virus recombinants de la vaccine ont été obtenus par double recombinaison homologue *in vivo* entre la cassette TK des plasmides de transfert des séries pTG et pTN et le gène TK de la souche Copenhague du virus de la vaccine selon une procédure décrite par Kieny et coll. (1984, Nature, 312 : 163-166).

5 Brièvement, des cellules CV-1 sont transfectées à l'aide de DOTAP (Roche) par de l'ADN génomique de la souche copenhague du virus de la vaccine et chacun des plasmides de transfert des séries pTG et pTN décrits ci-dessus, puis surinfectées par le virus vaccine auxiliaire VV-ts7 pendant 24 heures à 33°C. Le virus auxiliaire est contre-sélectionné par incubation à 40°C pendant 2 jours puis les virus recombinants
10 (phénotype TK-) sélectionnés par deux cycles de clonage sous milieu gélosé sur cellules 143Btk- en présence de BuDr (25 µg/ml). Les 6 virus VV-TG, VV-TG-S, VV-TG-Ssol, VV-TN, VV-TN-S et VV-TN-Ssol ont été respectivement obtenus à l'aide des plasmides de transfert pTG186poly, pTG-S, pTG-Ssol, pTN480, pTN-S, pTN-Ssol. Les virus VV-TG et VV-TN n'expriment aucun gène hétérologue et ont été
15 utilisés comme contrôle TK- dans les expériences. Les préparations de virus recombinants ont été réalisés sur monocouches de cellules CV-1 ou BHK-21 et le titre en unités formant plaque (u.f.p.) déterminé sur cellules CV-1 selon Earl et Moss (1998, Current Protocols in Molecular Biology, 16.16.1-16.16.13).

3) Caractérisation des virus recombinants

20 L'expression des transgènes codant la protéine S et le polypeptide Ssol a été recherchée par western blot.

Des monocouches de cellules CV-1 ont été infectées à une multiplicité de 2 par les différents virus vaccine recombinants VV-TG, VV-TG-S, VV-TG-Ssol, VV-TN, VV-TN-S et VV-TN-Ssol. Après 18 heures d'incubation à
25 37°C et sous 5% de CO₂, des extraits cellulaires ont été préparés en tampon de dépôt selon Laemmli, séparés sur un gel SDS à 8% de polyacrylamide puis transférés sur une membrane de PVDF (BioRad). La détection de cette immunoempreinte (« western blot ») a été réalisée à l'aide d'un sérum polyclonal de lapin anti-S (immun-sérum du lapin P11135 : cf exemple 4) et d'anticorps polyclonaux d'âne dirigés contre les IgG
30 de lapin et couplés à la peroxidase (NA934V, Amersham). Les anticorps fixés ont été révélés par luminescence à l'aide du kit ECL+ (Amersham) et de films d'autoradiographie Hyperfilm MP (Amersham).

Comme le montre la figure 35A, le virus recombinant VV-TN-S dirige l'expression de la protéine S à des niveaux qui sont comparables à ceux que l'on peut observer 8h après infection par le SRAS-CoV mais qui sont bien plus élevés que ceux que l'on peut observer après infection par VV-TG-S. Dans une deuxième
5 expérience (figure 35B), l'analyse de quantités variables d'extraits cellulaires montre que les niveaux d'expression observés après infection par les virus de la série TN (VV-TN-S et VV-TN-Ssol) sont environ 10 fois plus élevés que ceux observés avec les virus de la série TG (VV-TG-S et VV-TG-Ssol respectivement). En outre, le polypeptide Ssol est sécrété dans le surnageant de cellules CV-1 infectées par le virus
10 VV-TN-Ssol plus efficacement que dans le surnageant de cellules infectées par VV-TG-Ssol (figure 36A). Dans cette expérience, le virus VV-TN-Sflag a été utilisé à titre de contrôle, car il exprime la forme membranaire de la protéine S fusionnée en son extrémité C-ter à l'étiquette FLAG. La protéine Sflag n'est pas détectée dans le surnageant des cellules infectées par VV-TN-Sflag, démontrant que le polypeptide
15 Ssol est bien sécrété de façon active après infection par VV-TN-Ssol.

Ces résultats démontrent que les virus vaccine recombinants sont bien porteurs des transgènes et permettent l'expression de la glycoprotéine du SRAS sous sa forme membranaire (S) ou sous une forme soluble et sécrétée (Ssol). Les virus vaccine porteurs du promoteur synthétique 480 permettent l'expression de S et la
20 sécrétion de Ssol à des niveaux bien plus élevés que les virus porteurs du promoteur du gène 7.5K.

4) Application à la production d'une forme soluble de la S du SRAS-CoV. Purification de cet antigène recombinant et applications diagnostiques.

La lignée BHK-21 est la lignée cellulaire qui sécrète les quantités les
25 plus élevées de polypeptide Ssol après infection par le virus VV-TN-Ssol parmi les lignées testées (BHK-21, CV1, 293T et FrhK-4, figure 36B); elle permet la production et la purification en quantité du polypeptide recombinant Ssol. Dans une expérience type où les conditions expérimentales d'infection, de production et de purification ont été optimisées, les cellules BHK-21 sontensemencées en milieu de
30 culture standard (DMEM sans pyruvate contenant 4.5g/l de glucose et supplémenté par 5% de TPB, 5% de SVF, 100 U/ml de pénicilline et 100 µg/ml de streptomycine) sous la forme d'une monocouche sous-confluente (10 millions de cellules pour chaque

100 cm² dans 25 ml de milieu). Après 24h d'incubation à 37°C sous 5% de CO₂, les cellules sont infectées à une M.O.I. de 0.03 et le milieu standard remplacé par le milieu de sécrétion où la quantité de SVF est abaissée à 0.5% et le TPB supprimé. Le surnageant de culture est prélevé après 2,5 jours d'incubation à 35°C et sous 5% de CO₂ et le virus de la vaccine inactivé par addition de triton X-100 (0,1%). Après filtration sur membrane de polyethersulfone (PES) de 0.1 µm, le polypeptide recombinant Ssol est purifié par une chromatographie d'affinité sur matrice anti-FLAG avec élution par une solution de peptide FLAG (DYKDDDDK) à 100 µg/ml en TBS (Tris 50mM pH 7.4, 150 mMNaCl).

10 L'analyse par gel SDS à 8 % d'acrylamide coloré au nitrate d'argent a mis en évidence un polypeptide majoritaire dont la masse moléculaire est d'environ 180kD et dont le degré de pureté est supérieur à 90% (figure 37). La concentration du polypeptide recombinant Ssol purifié a été déterminée par comparaison avec les marqueurs de masse moléculaire et estimée à 24 ng/µl.

15 Cette préparation de polypeptide Ssol purifié permet de réaliser une gamme étalon pour mesurer à l'aide d'un test ELISA-capture les concentrations de Ssol présentes dans les surnageants de culture. Selon ce test, la lignée BHK-21 sécrète environ 1 µg/ml de polypeptide Ssol dans les conditions de production décrites plus haut. En outre, le schéma de purification présenté permet de purifier de l'ordre de 160 µg de polypeptide Ssol par litre de surnageant de culture.

La réactivité en ELISA du polypeptide recombinant Ssol a été analysée vis-à-vis de sérums de patients atteints de SRAS.

25 Les sérums de cas probables de SRAS testés ont été choisis sur la base des résultats (positifs ou négatifs) d'analyse de leur réactivité spécifique vis à vis des antigènes natifs du SRAS-CoV par test d'immunofluorescence sur cellules VeroE6 infectées par le SRAS-CoV et/ou par test ELISA indirect en utilisant comme antigène un lysat de cellules VeroE6 infectées par le SRAS-CoV. Les sérums de ces patients sont identifiés par un numéro d'ordre du Centre National de Référence des Virus Influenzae ainsi que par les initiales du patient et le nombre de jours écoulés depuis le début des symptômes. Tous les sérums de cas probables (cf Tableau XVI)

30 reconnaissent les antigènes natifs du SRAS-CoV à l'exception du sérum 032552 du patient VTT, pour lequel l'infection par le SRAS-CoV n'a pas pu être confirmée par

RT-PCR réalisée sur prélèvements respiratoires des jours 3, 8 et 12. Un panel de sérums témoin a été utilisé à titre de contrôle (sérums TV): il s'agit de sérums prélevés en France avant l'épidémie de SRAS survenue en 2003.

5 **Tableau XVI : sérums de cas probables de SRAS**

sérum	patient	jour de prélèvement
033168	JYK	38
033597	JYK	74
032632	NTM	17
032634	THA	15
032541	PHV	10
032542	NIH	17
032552	VTT	8
032633	PTU	16

Des phases solides sensibilisées par le polypeptide recombinant Ssol ont été préparées par adsorption d'une solution de polypeptide Ssol purifié à 4 µg/ml en PBS dans les puits d'une plaque ELISA. Les plaques sont incubées une nuit à 4°C puis lavées avec du tampon PBS-Tween (PBS, 0,1% Tween20). Après lavage en PBS-tween, les sérums à tester (100 µl) sont dilués au 1/100 et 1/400 dans du tampon PBS-lait écrémé-Tween (PBS, 3% lait écrémé, 0,1% Tween) puis ajoutés dans les puits de la plaque ELISA sensibilisée. Les plaques sont incubées 1h à 37°C. Après 3 lavages avec du tampon PBS-Tween, le conjugué anti-IgG humaines marqué à la peroxidase (réf NA933V, Amersham) dilué au 1/4000 dans du tampon PBS-lait écrémé-Tween est ajouté puis les plaques sont incubées une heure à 37°C. Après 6 lavages avec du tampon PBS-Tween, le chromogène (TMB) et le substrat (H₂O₂) sont ajoutés et les plaques sont incubées 10 minutes à l'abri de la lumière. La réaction est arrêtée par ajout d'une solution 1M de H₃PO₄ puis l'absorbance est mesurée à 450 nm avec une référence à 620nm.

Les tests ELISA (figure 38) démontrent que le polypeptide recombinant Ssol est reconnu spécifiquement par les anticorps sériques de patients atteints de SRAS prélevés en phase moyenne ou tardive de l'infection (≥ 10 jours après le début des symptômes), alors qu'il n'est pas reconnu de façon significative par les anticorps sériques des sérums témoins de sujets non atteints de SRAS.

En conclusion, ces résultats démontrent que le polypeptide recombinant Ssol peut être purifié à partir du surnageant de cellules de mammifères infectées par le virus vaccine recombinant VV-TN-Ssol et être utilisé comme antigène pour la mise au point d'un test ELISA de diagnostic sérologique de l'infection par le
5 SRAS-CoV.

5) Applications vaccinales

L'immunogénicité des virus vaccine recombinants a été étudiée chez la souris.

Pour cela, des groupes de 7 souris BALB/c ont été immunisés par
10 voie i.v. à deux reprises à 4 semaines d'intervalle par 10^6 u.f.p. de virus vaccine recombinants VV-TG, VV-TG-S, VV-TG-Ssol, VV-TN, VV-TN-S et VV-TN-Ssol ainsi que, à titre de contrôle, VV-TG-HA qui dirige l'expression de l'hémagglutinine de la souche A/PR/8/34 du virus de la grippe. Les sérums immuns ont été prélevés 3 semaines après chacune des immunisations (IS1, IS2).

15 Les sérums immuns ont été analysés par pool pour chacun des groupes par ELISA indirect en utilisant un lysat de cellules VeroE6 infectées par le SRAS-CoV comme antigène et à titre de contrôle, un lysat de cellules VeroE6 non infectées. Les titres (TI) en anticorps anti-SRAS-CoV sont calculés comme l'inverse de la dilution produisant une DO spécifique de 0,5 après révélation par un anticorps
20 polyclonal anti-IgG(H+L) de souris couplé à la peroxidase (NA931V, Amersham) et du TMB additionné de H_2O_2 (KPL). Cette analyse (figure 39A) montre que l'immunisation par le virus VV-TG-S et VV-TN-S induit chez la souris dès la première immunisation des anticorps dirigés contre la forme native de la protéine de spicule du SRAS-CoV présente dans le lysat de cellules VeroE6 infectées. Les
25 réponses induites par le virus VV-TN-S sont plus élevées que celles induites par le virus VV-TG-S après la première (TI=740 et TI=270 respectivement) et la deuxième (TI=3230 et TI=600 respectivement) immunisation. Le virus VV-TN-Ssol induit de forts titres d'anticorps anti-SRAS-CoV après deux immunisations (TI=640), alors que le virus VV-TG-Ssol induit une réponse à la limite de la détection (TI=40).

30 Les sérums immuns ont été analysés par pool pour chacun des groupes pour leur capacité à séroneutraliser l'infectivité du SRAS-CoV. 4 points de séroneutralisation sur cellules FRhK-4 (100 TCID₅₀ de SRAS-CoV) sont réalisés

pour chacune des dilutions de raison 2 testées à partir du 1/20. Le titre séroneutralisant est calculé selon la méthode de Reed et Munsch comme l'inverse de la dilution neutralisant l'infectivité de 2 cupules sur 4. Cette analyse montre que les anticorps induits chez la souris par les virus vaccine exprimant la protéine S ou le polypeptide Ssol sont neutralisants et que les virus à promoteurs synthétiques sont des immunogènes plus efficaces que les virus porteurs du promoteur 7.5K : les titres les plus élevés (640) sont observés après 2 immunisations par le virus VV-TN-S (figure 39B).

Le pouvoir protecteur des anticorps neutralisants induits chez la souris après immunisation par les virus vaccine recombinants est évalué à l'aide d'une infection d'épreuve par le SRAS-CoV.

6) Autres applications

Des virus vaccine recombinants de troisième génération sont construits en substituant les séquences sauvages des gènes S et Ssol par des gènes synthétiques optimisés pour l'expression en cellules de mammifères, décrits ci-dessus. Ces virus vaccine recombinants sont susceptibles d'exprimer des quantités plus importantes des antigènes S et Ssol et donc de présenter une immunogénicité accrue.

Le virus vaccine recombinant VV-TN-Ssol peut être utilisé pour la production en quantité et la purification de l'antigène Ssol en vue d'applications diagnostiques (sérologie par ELISA) et vaccinales (vaccin sous-unitaire).

Exemple 17 : Virus recombinant de la rougeole exprimant la protéine de spicule (S) du coronavirus associé au SRAS (SRAS-CoV). Applications vaccinales.

1) Introduction

Le vaccin rougeole (MV) induit chez l'homme une immunité protectrice de longue durée après une seule injection (Hilleman, 2002, Vaccine, 20 : 651-665). La protection conférée est très robuste et repose sur l'induction d'une réponse en anticorps et d'une réponse cellulaire CD4 et CD8. Le génome du MV est très stable et aucune réversion vers la virulence des souches vaccinales n'a jamais été observée. Le virus de la rougeole appartient au genre des *Morbillivirus* de la famille des *Paramyxoviridae* ; c'est un virus enveloppé dont le génome est un ARN monocaténaire de polarité négative de 16kb (figure 40A) et dont le cycle de réplication exclusivement cytoplasmique exclut toute possibilité d'intégration dans le

génomique de l'hôte. Le vaccin rougeole est ainsi l'un des vaccins vivants les plus efficaces et les plus sûrs utilisés dans la population humaine. L'équipe de Frédéric Tangy a développé récemment un vecteur d'expression sur la base de la souche Schwarz du virus de la rougeole, qui est la souche atténuée la plus sûre et la plus
5 utilisée chez l'homme comme vaccin contre la rougeole. Cette souche vaccinale peut être isolée à partir d'un clone moléculaire infectieux en conservant son immunogénicité chez les primates ainsi que chez la souris susceptible à l'infection. Elle constitue, après insertion d'unités de transcription supplémentaires, un vecteur pour l'expression de séquences hétérologues (Combredet, 2003, J. Virol. 77 : 11546-
10 11554). En outre, un MV Schwarz recombinant exprimant la glycoprotéine d'enveloppe du virus West Nile (WNV) induit une réponse en anticorps efficace et de longue durée qui protège la souris d'une infection d'épreuve létale par le WNV (Despres et al, 2004, J. Infect. Dis., sous presse). Toutes ces caractéristiques font de la souche atténuée Schwarz du virus de la rougeole un candidat vecteur extrêmement
15 prometteur pour la construction de nouveaux vaccins vivants recombinants.

Le but de cet exemple est d'évaluer la capacité de virus recombinants de la rougeole (MV) exprimant différents antigènes du coronavirus associé au SRAS (SRAS-CoV) à constituer de nouveaux candidats vaccins contre le SRAS.

20 Les inventeurs se sont intéressés à la protéine de spicule (S) du SRAS-CoV, qui permet d'induire après immunisation génique chez l'animal des anticorps neutralisants l'infectivité du SRAS-CoV, ainsi qu'à une forme soluble et sécrétée de cette protéine, le polypeptide Ssol, qui est composé de l'ectodomaine (aa 1-1193) de la S fusionné en son extrémité C-ter à une étiquette FLAG (DYKDDDDK)
25 via un linker BspE1 codant pour le dipeptide SG. Ce polypeptide Ssol présente une antigénicité similaire à celle de la protéine S et permet, après injection à la souris sous la forme d'une protéine purifiée adjuvée en hydroxide d'aluminium, l'induction de titres élevés d'anticorps neutralisants contre le SRAS-CoV.

Les différentes formes du gène S ont été introduites sous la forme
30 d'une unité de transcription supplémentaire entre les gènes P (phosphoprotéine) et M (matrice) dans le cADN de la souche Schwarz du MV précédemment décrite (Combredet, 2003, J. Virol. 77 : 11546-11554 ; Demande EP N° 02291551.6 du 20

juin 2002, et Demande EP n° 02291550.8 du 20 juin 2002). Après avoir isolé les virus recombinants MVSchw2-SARS-S et MVSchw2-SARS-Ssol et vérifié leur capacité à exprimer l'antigène S du SRAS-CoV, leur capacité à induire chez la souris puis chez le singe une réponse immunitaire protectrice contre le SRAS est testée.

5 2) Construction des virus recombinants

Le plasmide pTM-MVSchw-ATU2 (figure 40B) contient un cADN infectieux correspondant à l'antigénome de la souche vaccinale Schwarz du virus de la rougeole (MV) dans lequel une unité de transcription supplémentaire (ATU) a été introduite entre les gènes P (phosphoprotéine) et M (matrice) (Combredet, 2003,
10 Journal of Virology, 77 : 11546-11554). Des génomes recombinants MVSchw2-SARS-S et MVSchw2-SARS-Ssol du virus de la rougeole ont été construits par l'insertion des ORFs de la protéine S et du polypeptide Ssol au sein de l'unité de transcription supplémentaire du vecteur MVSchw-ATU2.

Pour cela, un fragment d'ADN contenant le cADN de la S du SRAS-
15 CoV a été amplifié par PCR à l'aide des oligonucléotides 5'-ATACGTACGA CCATGTTTAT TTTCTTATTA TTTCTTACTC TCACT-3' et 5'-ATAGCGCGCT CATTATGTGT AATGTAATTT GACACCCTTG-3' en utilisant le plasmide pcDNA-S comme matrice puis inséré dans le plasmide pCR®2.1-TOPO (Invitrogen) pour obtenir le plasmide pTOPO-S-MV. Les deux oligonucléotides utilisés
20 contiennent des sites de restriction BsiW1 et BssHII, de façon à permettre l'insertion ultérieure dans le vecteur rougeole, et ont été conçus de façon à générer une séquence de 3774 nt incluant les codons d'initiation et de terminaison, afin de respecter la règle des 6 qui stipule que la longueur du génome d'un virus rougeole doit être divisible par 6 (Calain & Roux, 1993, J. Virol., 67 : 4822-4830 ; Schneider et al., 1997, Virology,
25 227 : 314-322). L'insert a été séquencé à l'aide d'un kit BigDye Terminator v1.1 (Applied Biosystems) et d'un séquenceur automatique ABI377.

Afin d'exprimer une forme soluble et sécrétée de la S du SRAS-CoV, un plasmide contenant l'ADNc du polypeptide Ssol correspondant à l'ectodomaine (aa 1-1193) de la S du SRAS-CoV fusionné en son extrémité C-ter à la
30 séquence d'une étiquette FLAG (DYKDDDDK) via un linker BspE1 codant pour le dipeptide SG a ensuite été obtenu. Pour cela, un fragment d'ADN a été amplifié à l'aide des oligonucléotides 5'-CCATTTC AAC AATTTGGCCG-3' et 5'-

ATAGGATCCG CGCGCTCATT ATTTATCGTC GTCATCTTTA TAATC-3' à partir du plasmide pCDNA-Ssol puis inséré dans le plasmide pTOPO-S-MV entre les sites SalI et BamHI pour obtenir le plasmide pTOPO-S-MV-SF. La séquence générée est longue de 3618 nt entre les sites BsiWI et BssHII et respecte la règle des 6.

5 L'insert a été séquencé comme indiqué ci-dessus.

Les fragments BsiWI-BssHII contenant les ADNc de la protéine S et du polypeptide Ssol ont ensuite été excisés par digestion des plasmides pTOPO-S-MV et pTOPO-S-MV-SF puis sous-clonés entre les sites correspondants du plasmide pTM-MVSchw-ATU2 pour donner les plasmides pTM-MVSchw2-SARS-S et pTM-MVSchw2-SARS-Ssol (figure 40B). Ces deux plasmides ont été déposés à la C.N.C.M. le 1^{er} décembre 2004, sous les numéros I-3326 et I-3327, respectivement.

Les virus rougeole recombinants correspondants aux plasmides pTM-MVSchw2-SARS-S et pTM-MVSchw2-SARS-Ssol ont été obtenus par génétique inverse selon le système reposant sur l'utilisation d'une lignée cellulaire auxiliaire, décrit par Radecke et coll. (1995, Embo J., 14, : 5773-5784) et modifié par Parks et coll. (1999, J. Virol., 73 : 3560-3566). Brièvement, les cellules auxiliaires 293-3-46 sont transfectées selon la méthode au phosphate de calcium par 5 µg des plasmides pTM-MVSchw2-SARS-S ou pTM-MVSchw2-SARS-Ssol et 0,02 µg du plasmide pEMC-La dirigeant l'expression de la polymérase L du MV (don de M.A. Billeter). Après une nuit d'incubation à 37°C, un choc thermique est réalisé pendant 2 heures à 43°C et les cellules transfectées sont transférées sur une monocouche de cellules Vero. Pour chacun des deux plasmides, des syncytia sont apparus après 2 à 3 jours de co-culture et ont été transférés successivement sur des monocouches de cellules Vero à 70% de confluence en boîtes de pétri de 35 mm puis en flacons de 25 et 75 cm². Quand les syncytia ont atteint 80-90% de confluence, les cellules sont récupérées à l'aide d'un grattoir puis congelées et décongelées une fois. Après une centrifugation à basse vitesse, le surnageant contenant le virus est conservé en aliquot à -80°C. Les titres des virus recombinants MVSchw2-SARS-S et MVSchw2-SARS-Ssol ont été déterminés par dilution limite sur cellules Vero et le titre en dose infectant 50% des cupules (TCID₅₀) calculé selon la méthode de Kärber.

3) Caractérisation des virus recombinants

L'expression des transgènes codant la protéine S et le polypeptide Ssol a été recherchée par western blot et immunofluorescence.

Des monocouches de cellules Vero en flacons T-25 ont été infectées à une multiplicité de 0,05 par différents passages des deux virus MVSchw2-SARS-S et MVSchw2-SARS-Ssol et le virus sauvage MWSchw à titre de contrôle. Quand les syncytia ont atteint 80 à 90% de confluence, des extraits cytoplasmiques ont été préparés dans un tampon d'extraction (150mMNaCl, 50 mM Tris-HCl pH 7,2, 1% triton-X-100, 0,1% SDS, 1% DOC) puis dilués en tampon de dépôt selon Laemmli, séparés sur un gel SDS à 8% de polyacrylamide et transférés sur une membrane de PVDF (BioRad). La détection de cette immunoempreinte (« western blot ») a été réalisée à l'aide d'un sérum polyclonal de lapin anti-S (immun-sérum du lapin P11135 : cf exemple 4 ci-dessus) et d'anticorps polyclonaux d'âne dirigés contre les IgG de lapin et couplés à la peroxidase (NA934V, Amersham). Les anticorps fixés ont été révélés par luminescence à l'aide du kit ECL+ (Amersham) et de films d'autoradiographie Hyperfilm MP (Amersham).

Des cellules Vero en monocouches sur lamelles de verre ont été infectées par les deux virus MVSchw2-SARS-S et MVSchw2-SARS-Ssol et le virus sauvage MWSchw à titre de contrôle à des multiplicités d'infection de 0,05. Quand les syncytia ont atteint 90 à 100% (virus MVSchw2-SARS-Ssol) ou 30 à 40% (MVSchw2-SARS-S, MWSchw) de confluence, les cellules ont été fixées dans une solution de PBS-PFA 4%, perméabilisées par une solution de PBS contenant 0,2% de triton puis marquées par des anticorps polyclonaux de lapins hyperimmunisés par des virions purifiés et inactivés du SRAS-CoV et par un conjugué d'anticorps de chèvre anti-IgG(H+L) de lapin couplé au FITC (Jackson).

Comme le montrent les figures 41 et 42, les virus recombinants MVSchw2-SARS-S et MVSchw2-SARS-Ssol dirigent l'expression de la protéine S et du polypeptide Ssol respectivement à des niveaux comparables à ceux que l'on peut observer 8h après infection par le SRAS-CoV. L'expression de ces polypeptides est stable après 3 passages des virus recombinants en culture cellulaire. Ces résultats démontrent que les virus rougeole recombinants sont bien porteurs des transgènes et permettent l'expression de la glycoprotéine du SRAS sous sa forme membranaire (S)

ou sous une forme soluble (Ssol). On s'attend à ce que le polypeptide Ssol soit sécrété des cellules infectées par le virus MVSchw2-SARS-Ssol comme c'est le cas lorsque ce même polypeptide est exprimé dans des cellules de mammifères après transfection transitoire des séquences correspondantes (cf exemple 11 ci-dessus).

5 4) Applications

Ayant montré que les virus MVSchw2-SARS-S et MVSchw2-SARS-Ssol permettent l'expression de la S du SRAS-CoV, leur capacité à induire une réponse immunitaire protectrice contre le SRAS-CoV chez la souris CD46^{+/-} IFN- $\alpha\beta$ R^{-/-}, qui est susceptible à l'infection par le MV, est évaluée. La réponse en anticorps des souris immunisées est évaluée par test ELISA contre les antigènes natifs du SRAS-CoV et pour leur capacité à neutraliser l'infectivité du SRAS-CoV *in vitro*, en utilisant les méthodologies décrites ci-dessus. Le pouvoir protecteur de la réponse sera évalué en mesurant la réduction de la charge virale pulmonaire 2 jours après une infection d'épreuve non létale par le SRAS-CoV.

Des virus rougeole recombinants de seconde génération sont construits en substituant les séquences sauvages des gènes S et Sol par des gènes synthétiques optimisés pour l'expression en cellules de mammifères, décrits à l'exemple 15 ci-dessus. Ces virus rougeole recombinants sont susceptibles d'exprimer des quantités plus importantes des antigènes S et Ssol et donc de présenter une immunogénicité accrue.

Alternativement, les gènes sauvages ou synthétiques codant pour la protéine S ou le polypeptide Ssol peuvent être insérés dans le vecteur rougeole MVSchw-ATU3 sous la forme d'une unité de transcription supplémentaire localisée entre les gènes H et L, puis les virus recombinants produits et caractérisés de façon similaire. Cette insertion est susceptible de générer des virus recombinants possédant des caractéristiques (multiplication du virus, niveau d'expression du transgène) différentes et possiblement une immunogénicité améliorée par rapport à ceux obtenus après insertion des transgènes entre les gènes P et N.

Le virus rougeole recombinant MVSchw2-SARS-Ssol peut être utilisé pour la production en quantité et la purification de l'antigène Ssol en vue d'applications diagnostiques et vaccinales.

Exemple 18 : Autres applications liées à la protéine S

a) Les vecteurs lentiviraux permettant l'expression de la S ou de Ssol (voire de fragments de la S) peuvent constituer un vaccin recombinant contre le SRAS-CoV, pour être utilisé en prophylaxie humaine et vétérinaire. Afin de démontrer la faisabilité d'un tel vaccin, l'immunogénicité des vecteurs lentiviraux recombinants TRIP-SD/SA-S-WPRE et TRIP-SD/SA-Ssol-WPRE est étudiée chez la souris.

b) Des anticorps monoclonaux sont produits à l'aide du polypeptide recombinant Ssol. D'après les résultats présentés à l'exemple 14 ci-dessus, ces anticorps ou du moins la majorité d'entre eux reconnaîtront la forme native de la S du SRAS-CoV et seront susceptibles d'applications diagnostiques et/ou prophylactiques.

c) Un test de sérologie du SRAS est mis au point avec le polypeptide Ssol utilisé comme antigène et la méthodologie du double épitope.

REVENDICATIONS

1°) Souche isolée ou purifiée de coronavirus humain associé au syndrome respiratoire aigu sévère, caractérisée en ce que son génome présente sous la
5 forme d'ADN complémentaire un codon sérine en position 23220-23222 du gène de la protéine S ou un codon glycine en position 25298-25300 du gène de l'ORF3, et un codon alanine en position 7918-7920 de l'ORF1a ou un codon sérine en position 26857-26859 du gène de la protéine M, lesdites positions étant indiquées en référence à la séquence Genbank AY274119.3.

10 2°) Souche isolée ou purifiée de coronavirus selon la revendication 1, caractérisée en ce que l'équivalent ADN de son génome présente une séquence correspondant à la séquence SEQ ID NO : 1.

3°) Polynucléotide isolé ou purifié, caractérisé en ce que sa séquence est celle du génome de la souche isolée de coronavirus selon la revendication 1 ou la
15 revendication 2.

4°) Polynucléotide isolé ou purifié selon la revendication 3, caractérisé en ce que sa séquence est la SEQ ID NO : 1.

5°) Paire d'amorces apte à amplifier un fragment de la séquence du génome d'un coronavirus associé au SRAS ou de son équivalent ADN, caractérisée en
20 ce qu'elle est sélectionnée dans le groupe constitué par :

- la paire d'amorces n° 1 correspondant respectivement aux positions 28507 à 28522 (amorce sens, SEQ ID NO : 60) et 28774 à 28759 (amorce anti-sens, SEQ ID NO : 61) de la séquence du polynucléotide selon la revendication 3 ou la revendication 4,

25 - la paire d'amorces n° 2 correspondant respectivement aux positions 28375 à 28390 (amorce sens, SEQ ID NO : 62) et 28702 à 28687 (amorce anti-sens, SEQ ID NO : 63) de la séquence du polynucléotide selon la revendication 3 ou la revendication 4, et

- la paire d'amorces constituée des amorces SEQ ID Nos 55 et 56.

30 6°) Sonde apte à détecter la présence du génome d'un coronavirus associé au SRAS ou d'un fragment de celui-ci, caractérisée en ce qu'elle est sélectionnée dans le groupe constitué par les fragments correspondant aux positions

suivantes de la séquence du polynucléotide selon la revendication 3 ou à la revendication 4: 28561 à 28586, 28588 à 28608, 28541 à 28563 et 28565 à 28589 (SEQ ID NO : 64 à 67).

5 7°) Vecteur de clonage et/ou d'expression recombinant, caractérisé en ce qu'il comprend un insert de séquence SEQ ID NO :38 et qu'il est compris dans une souche bactérienne qui et qu'il a été déposée sous le n° I-3048, le 5 juin 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15.

10 8°) Vecteur de clonage et/ou d'expression recombinant, caractérisé en ce qu'il contient un fragment d'ADNc sélectionné dans le groupe constitué par :
- un fragment d'ADNc codant pour une fusion C-terminale de la protéine N (SEQ ID NO : 37) avec une étiquette polyhistidine, et
- un fragment d'ADNc codant pour une fusion N-terminale de la protéine N (SEQ ID NO : 37) avec une étiquette polyhistidine.

15 9°) Vecteur d'expression recombinant selon la revendication 8, caractérisé en ce qu'il est compris dans une souche bactérienne qui a été déposée sous le n° I- 3117, le 23 octobre 2003, auprès de la Collection Nationale de Cultures de Microorganismes, 25 rue du Docteur Roux, 75724 Paris Cedex 15.

20 10°) Cellules modifiées par un vecteur selon l'une quelconque des revendications 73 à 92.

11°) Hybridome produisant un anticorps monoclonal contre la protéine N, caractérisé en ce qu'il est choisi parmi les hybridomes suivants :
- l'hybridome produisant l'anticorps monoclonal 87, déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-3328,
25 - l'hybridome produisant l'anticorps monoclonal 86, déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-3329,
- l'hybridome produisant l'anticorps monoclonal 57, déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-3330, et
- l'hybridome produisant l'anticorps monoclonal 156, déposé à la
30 CNCM le 1^{er} décembre 2004 sous le numéro I-3331.

12°) Anticorps ou fragment d'anticorps polyclonal ou monoclonal dirigé contre la protéine N, caractérisé en ce qu'il est produit par un hybridome selon la revendication 11.

13°) Puce ou filtre, caractérisée en ce qu'il comprend un anticorps
5 ou un fragment d'anticorps selon la revendication 12.

14°) Test d'immunocapture destiné à détecter une infection par le coronavirus associé au SRAS, caractérisé en ce qu'il utilise un anticorps monoclonal spécifique de la nucléoprotéine virale native (protéine N).

15°) Test d'immunocapture selon la revendication 14, caractérisé en
10 ce que l'anticorps utilisé pour la capture de la nucléoprotéine virale native est un anticorps monoclonal spécifique de la région centrale et/ou d'un épitope conformationnel.

16°) Test d'immunocapture selon la revendication 14 ou 15, caractérisé en ce que l'anticorps utilisé pour la capture de la protéine N est l'anticorps
15 monoclonal Acm87, produit par l'hybridome déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-3328.

17°) Test d'immunocapture selon la revendication 14 ou 15, caractérisé en ce que l'anticorps utilisé pour la capture de la protéine N est l'anticorps monoclonal Acm86, produit par l'hybridome déposé à la CNCM le 1^{er} décembre 2004
20 sous le numéro I-3329.

18°) Test d'immunocapture selon la revendication 14 ou 15, caractérisé en ce que les anticorps monoclonaux Acm86 et Acm87 sont utilisés pour la capture de la protéine N.

19°) Test d'immunocapture selon l'une quelconque des
25 revendications 14 à 18, caractérisé en ce que l'anticorps utilisé pour la révélation de la protéine N est l'anticorps monoclonal Acm57, produit par l'hybridome déposé à la CNCM le 1^{er} décembre 2004 sous le numéro I-3330, ledit anticorps étant conjugué à une molécule ou une particule révélatrice.

20°) Test d'immunocapture selon l'une quelconque des
30 revendications 14 à 18, caractérisé en ce qu'une combinaison des anticorps Acm57 et Acm87, conjugués à une molécule ou une particule révélatrice, est utilisée pour la révélation de la protéine N.

21°) Réactif de détection d'un coronavirus associé au SRAS, caractérisé en ce qu'il est sélectionné dans le groupe constitué par :

(a) une paire d'amorces selon la revendication 5, ou une sonde selon la revendication 6,

5 (b) un vecteur recombinant selon l'une quelconque des revendications 7 à 9 ou une cellule modifiée selon la revendication 10,

(c) une souche isolée de coronavirus selon la revendication 1 ou la revendication 2 ou un polynucléotide selon l'une quelconque des revendications 3 et 4,

10 (d) un anticorps ou un fragment d'anticorps selon la revendication 12,

(e) une combinaison d'anticorps comprenant les anticorps monoclonaux Acm86 et/ou Acm87, et l'anticorps monoclonal Acm57;

(f) une puce ou un filtre selon la revendication 13.

15 22°) Utilisation d'un produit sélectionné dans le groupe constitué par : une paire d'amorces selon la revendication 50, une sonde selon la revendication 61, un vecteur recombinant selon l'une quelconque des revendications 7 à 913 à 22, une cellule modifiée selon la revendication 1023, une souche isolée de coronavirus selon la revendication 1 ou la revendication 2, un polynucléotide selon la a
20 revendication 3 ou la revendication ou 4, pour la préparation d'un réactif de détection et éventuellement de génotypage d'un coronavirus associé au SRAS.

23°) Méthode de détection d'un coronavirus associé au SRAS, à partir d'un échantillon biologique, laquelle méthode est caractérisée en ce qu'elle comprend au moins :

25 (a) l'extraction d'acides nucléiques présents dans ledit échantillon biologique,

(b) l'amplification d'un fragment de l'ORF-N par RT-PCR à l'aide d'un couple d'amorces selon la revendication 50, et

30 (c) la détection par tout moyen approprié des produits d'amplification obtenus en (b).

24°) Méthode selon la revendication 23, caractérisée en ce que l'étape (b) de détection est réalisée à l'aide d'au moins une sonde correspondant aux

positions 28561 à 28586, 28588 à 28608, 28541 à 28563 et 28565 à 28589 de la séquence du polynucléotide selon la revendication 3 ou la revendication 4.

25°) Méthode de détection d'une infection par un coronavirus associé au SRAS, à partir d'un échantillon biologique, par ELISA IgG indirect
5 utilisant la protéine N, laquelle méthode est caractérisée en ce que les plaques sont sensibilisées par une solution de protéine N à une concentration comprise entre 0,5 et 4µg/mL, de préférence 2µg/mL, dans un tampon PBS 10mM pH 7,2, rouge de phénol à 0,25mL /L.

26°) Méthode de détection d'une infection par un coronavirus
10 associé au SRAS, à partir d'un échantillon biologique, par ELISA double épitope, caractérisée en ce que le sérum à tester est mélangé à l'antigène de révélation, ledit mélange étant ensuite mis au contact de l'antigène fixé sur un support solide.

27°) Complexe immunitaire formé d'un anticorps ou d'un fragment
15 d'anticorps polyclonal ou monoclonal selon la revendication 11, et d'une protéine ou d'un peptide du coronavirus associé au SRAS.

28°) Kit ou coffret de détection d'un coronavirus associé au SRAS, caractérisé en ce qu'il comprend au moins un réactif sélectionné dans le groupe constitué par : une paire d'amorces selon la revendication 5, une sonde selon la revendication 6, un vecteur recombinant selon l'une quelconque des revendications 7
20 à 9, une cellule modifiée selon la revendication 10, une souche isolée de coronavirus selon la revendication 1 ou la revendication 2 et un polynucléotide selon la revendication 3 ou la revendication 4.

29°) Fragment du polynucléotide selon la revendication 3, caractérisé en ce qu'il inclut au moins un couple de bases ou de paires de bases
25 correspondant aux positions suivantes : 7919 et 23220, 7919 et 25298, 16622 et 23220, 19064 et 23220, 16622 et 25298, 19064 et 25298, 23220 et 24872, 23220 et 26857, 24872 et 25298, 25298 et 26857.

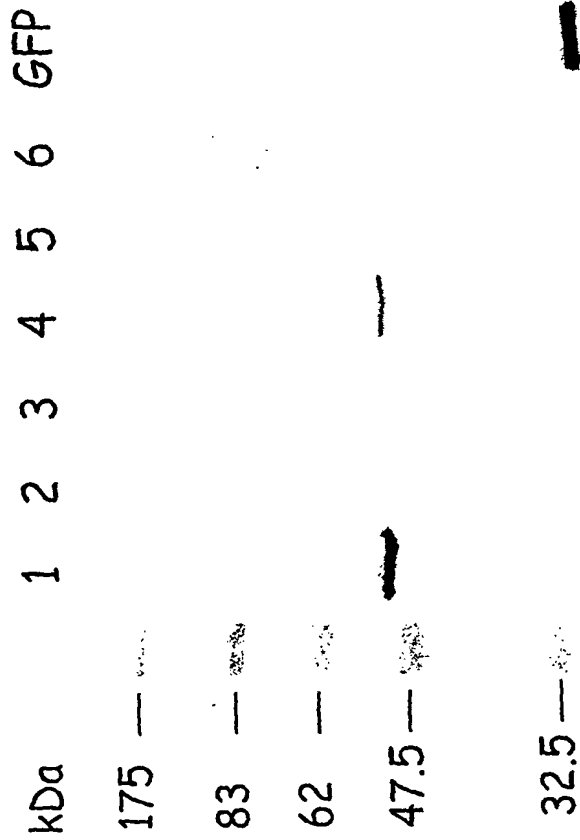


FIGURE 1

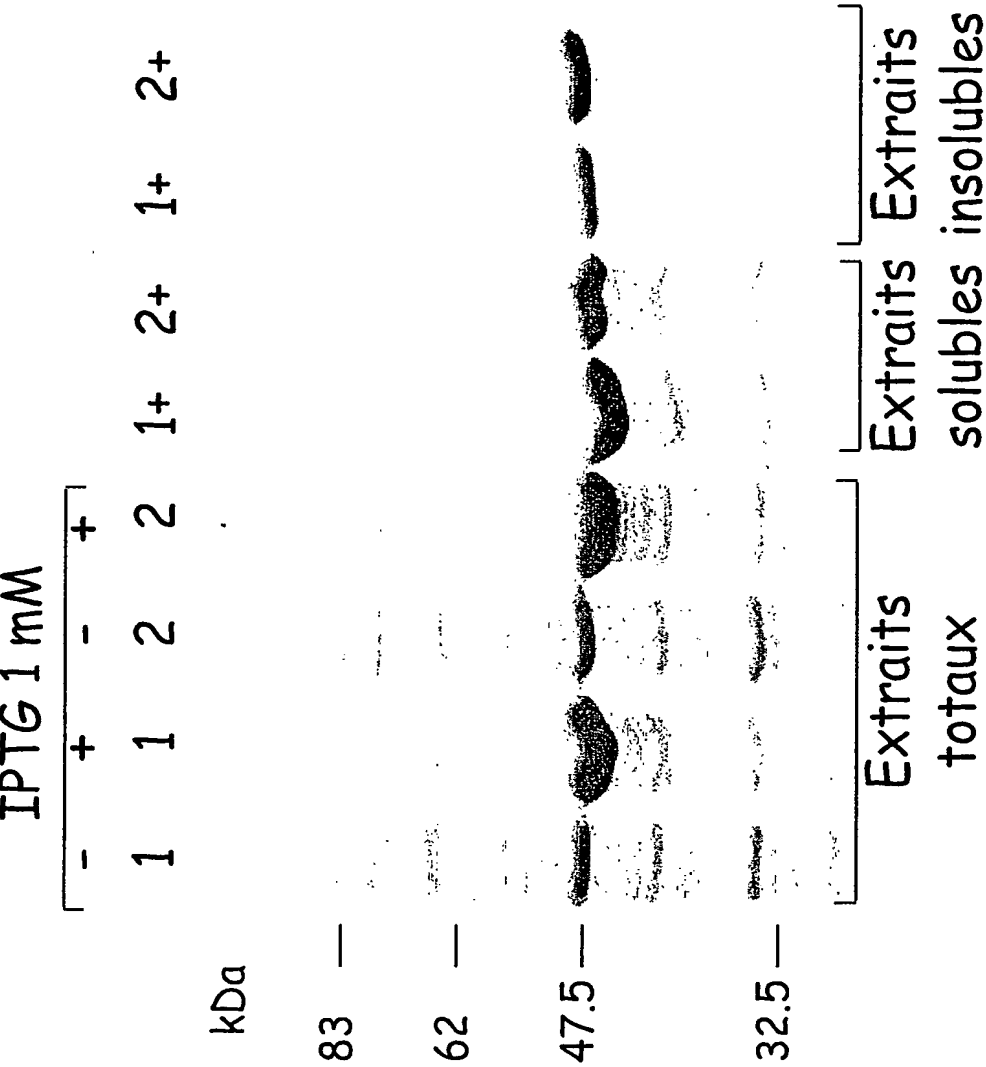


FIGURE 2

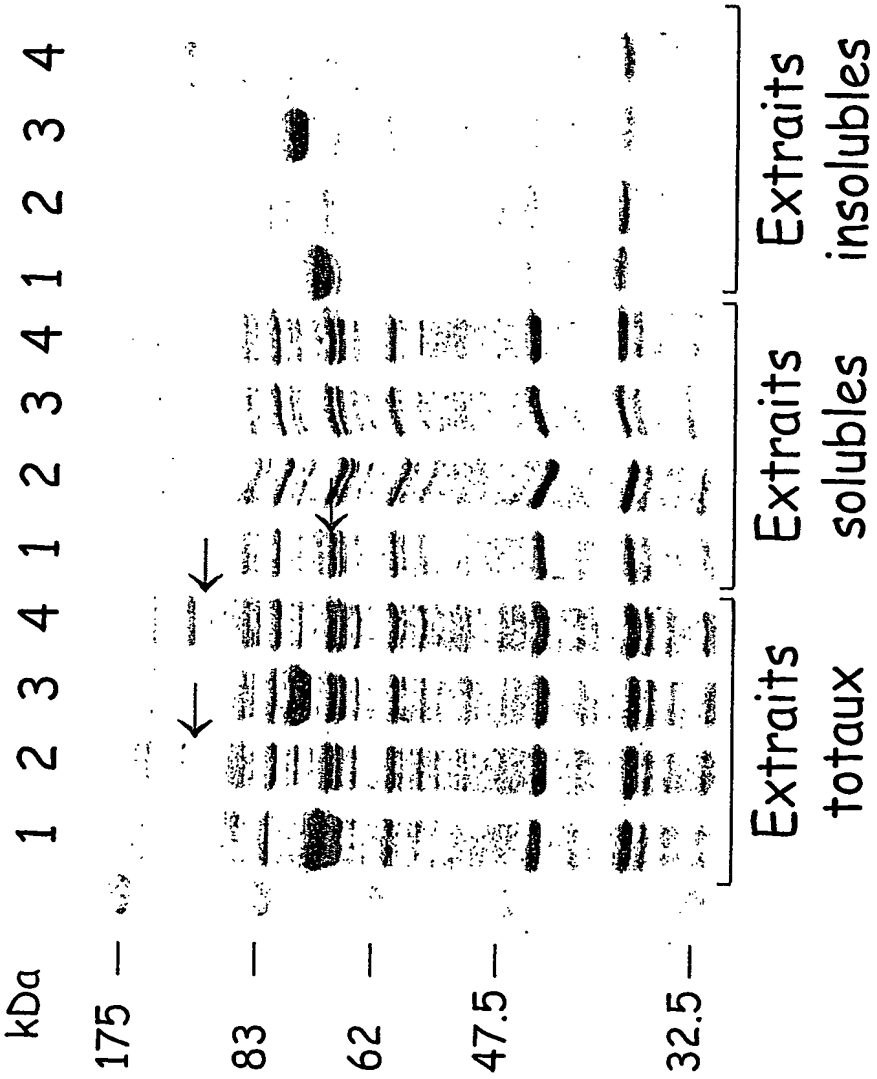


FIGURE 3

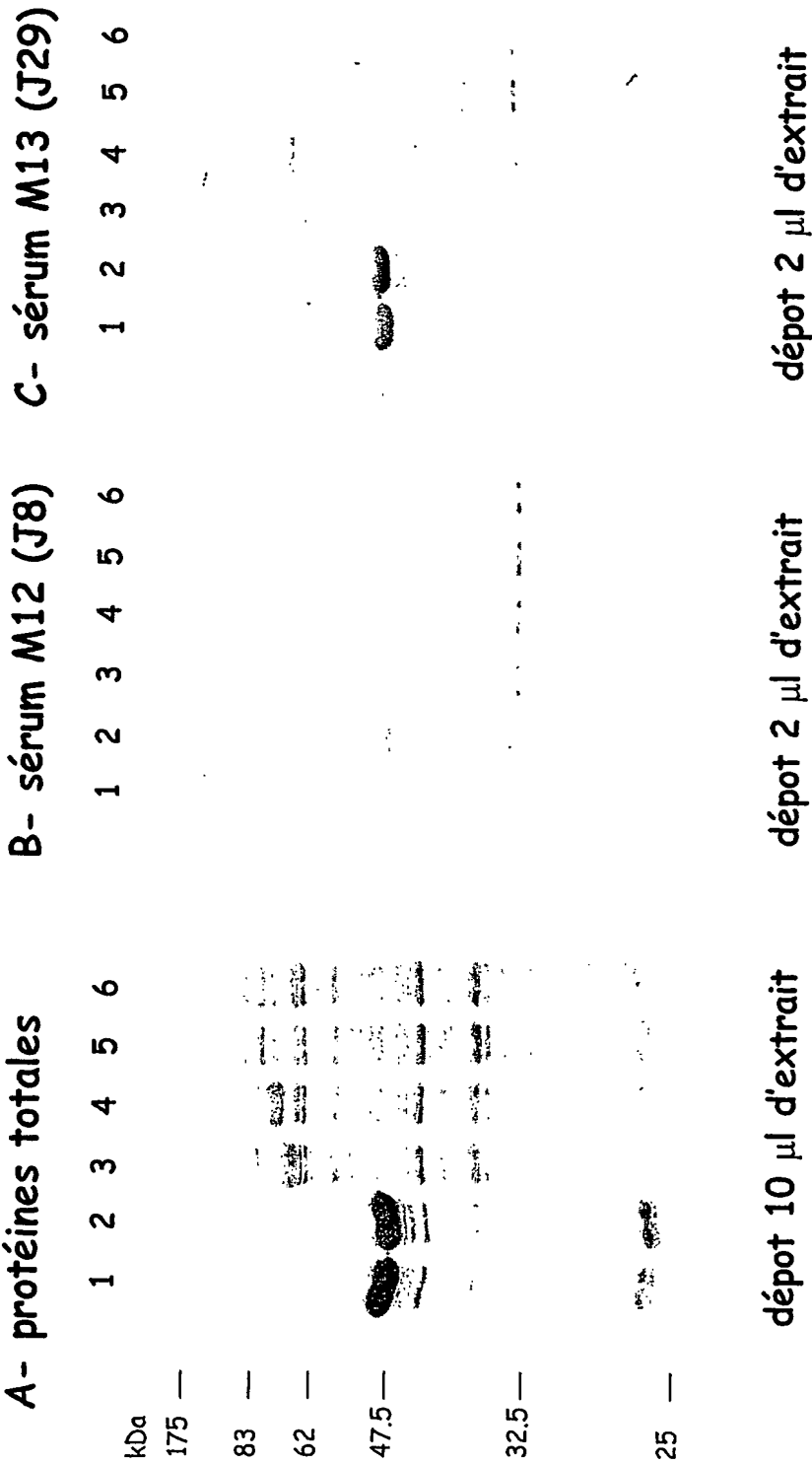


FIGURE 4

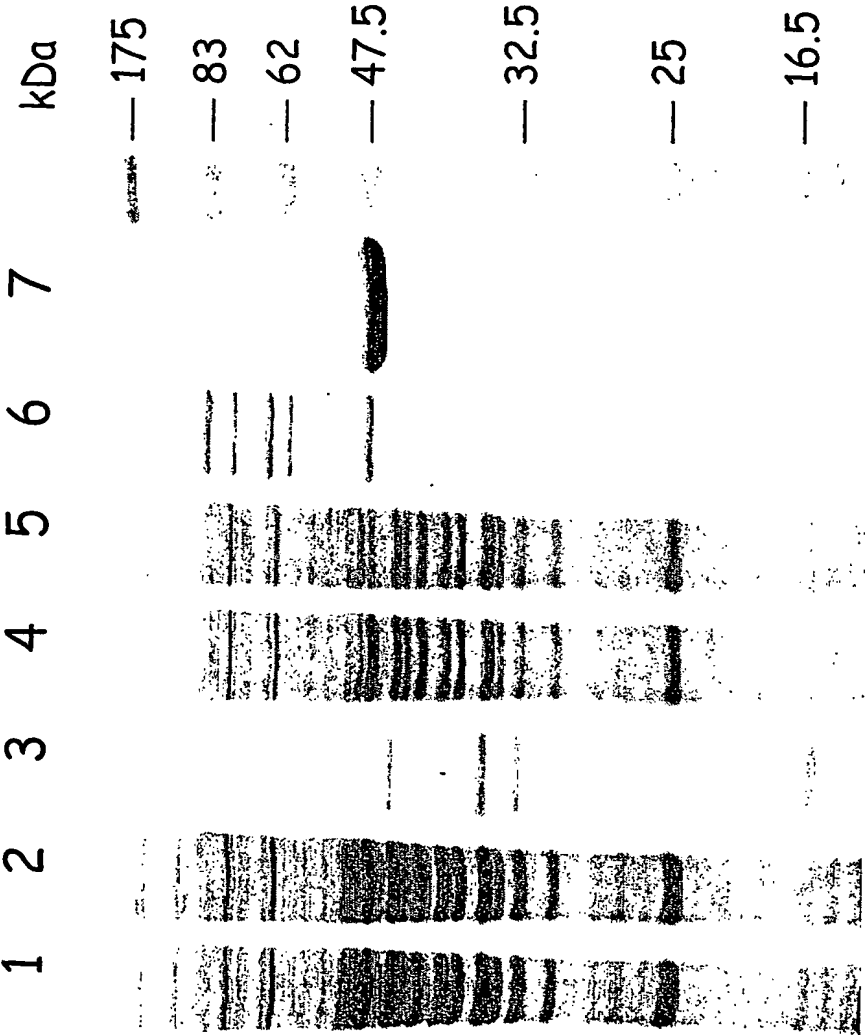


FIGURE 5

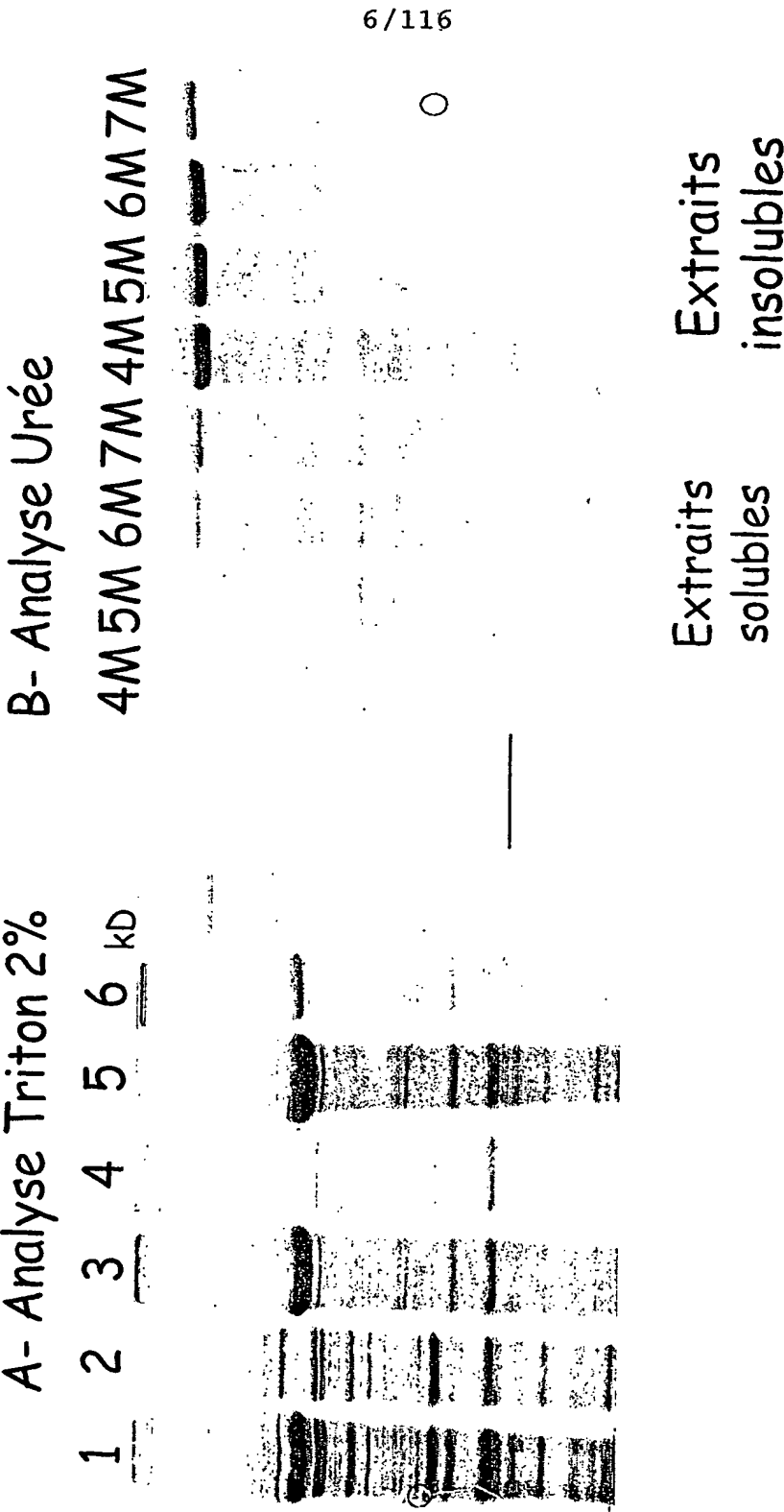


FIGURE 6

7/116

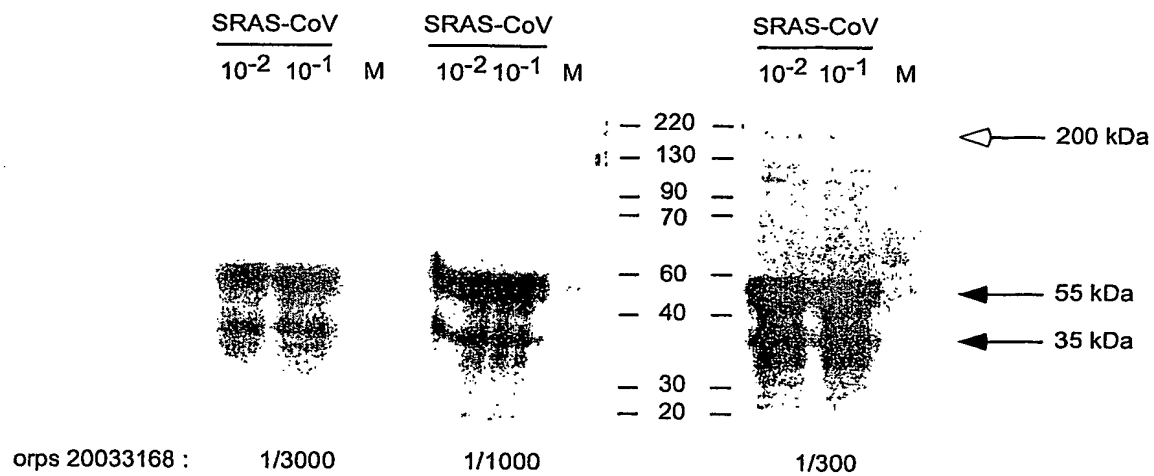


FIGURE 7

8/116

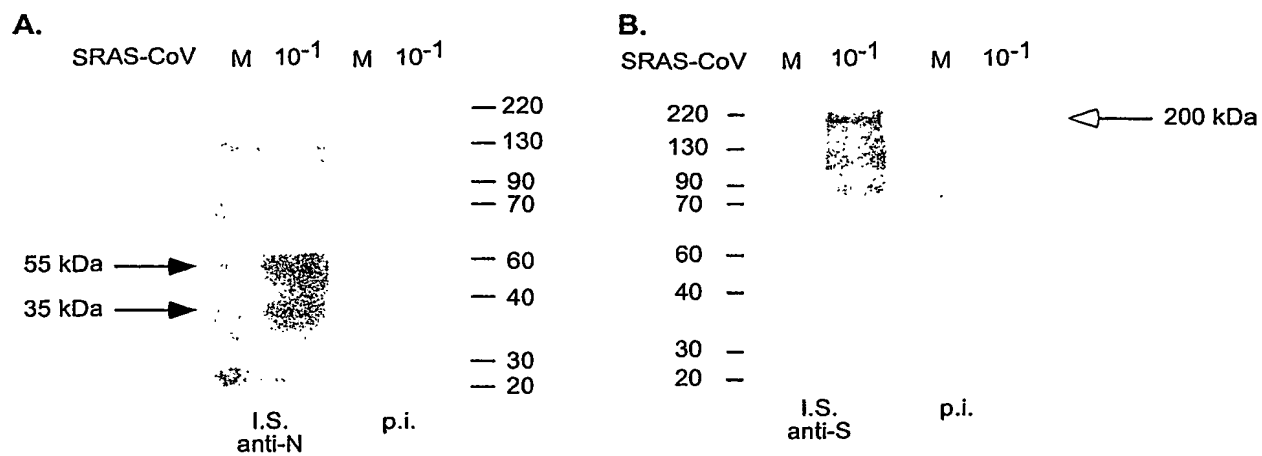
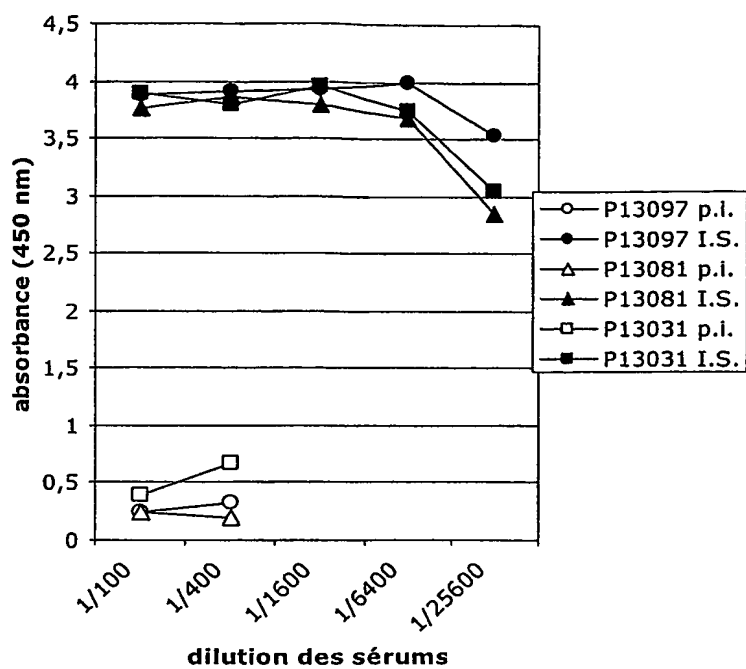


FIGURE 8

9/116

A



B

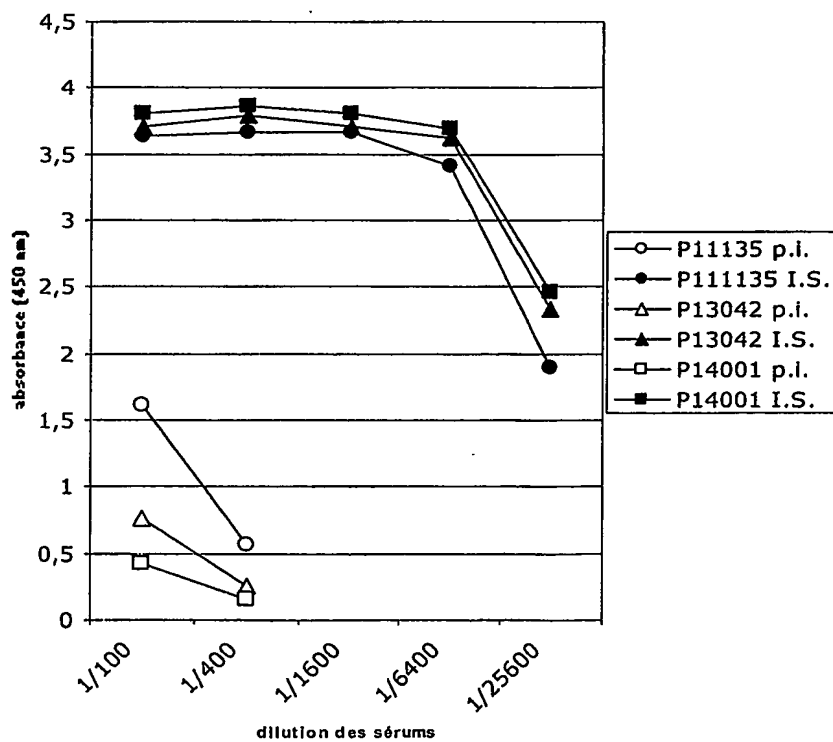


FIGURE 9

10/116

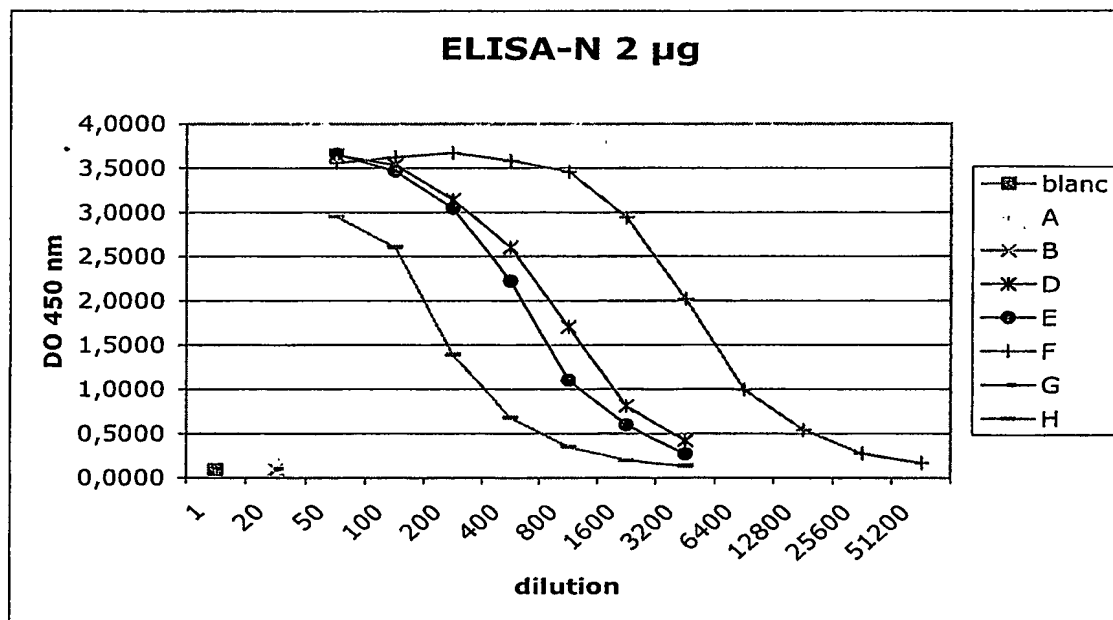
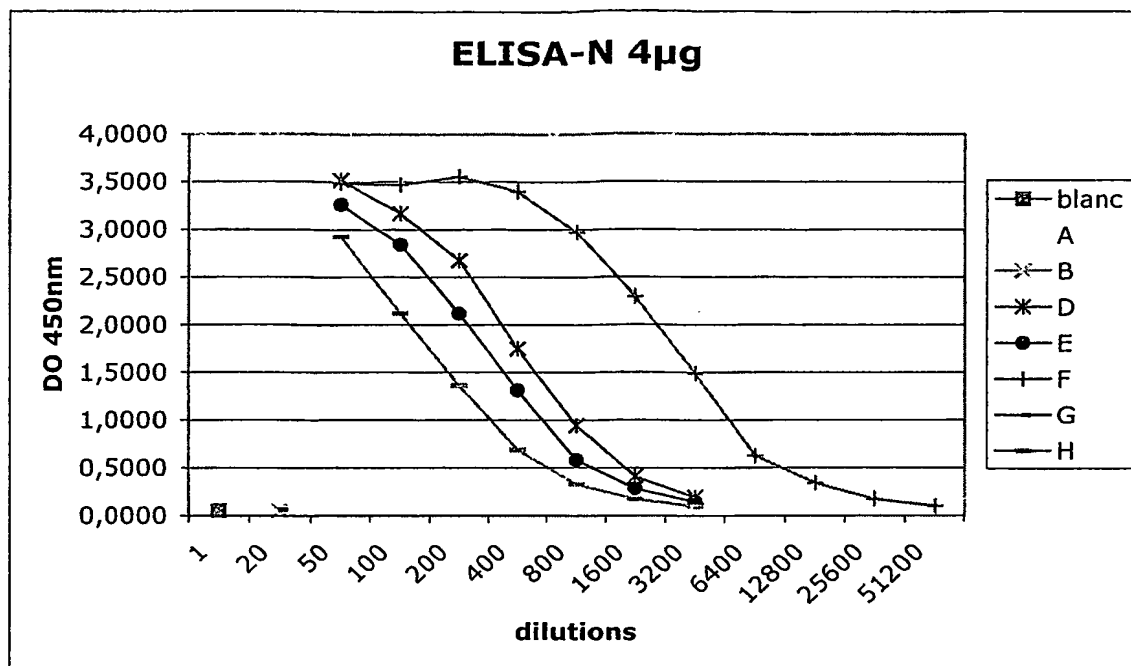


FIGURE 10a

11/116

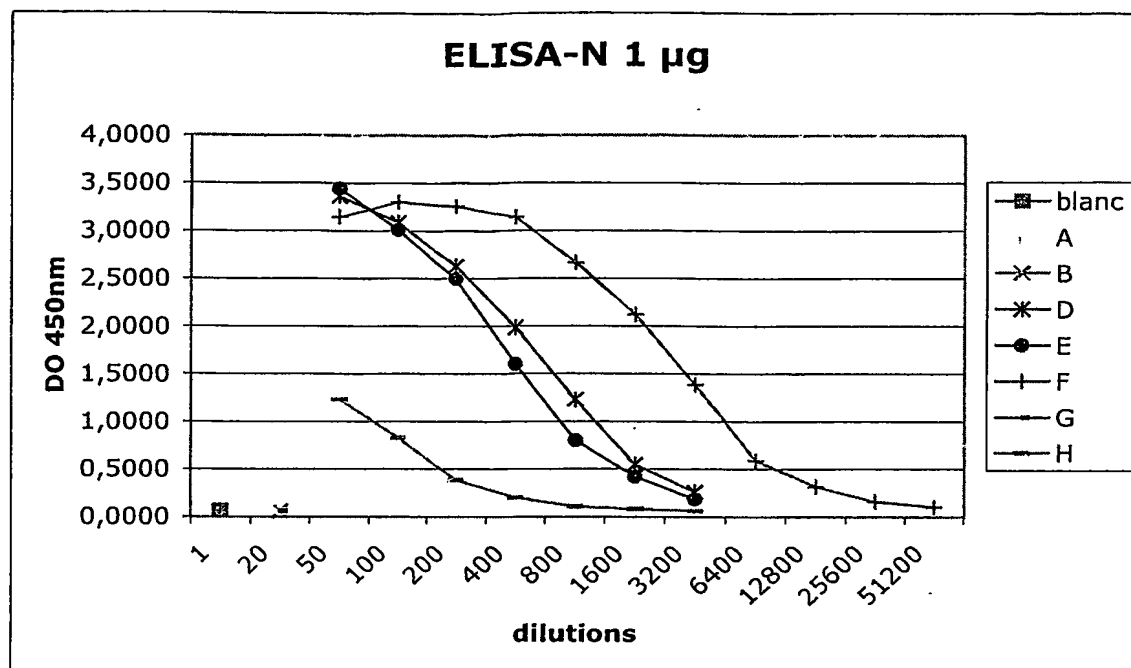


FIGURE 10b

12/116

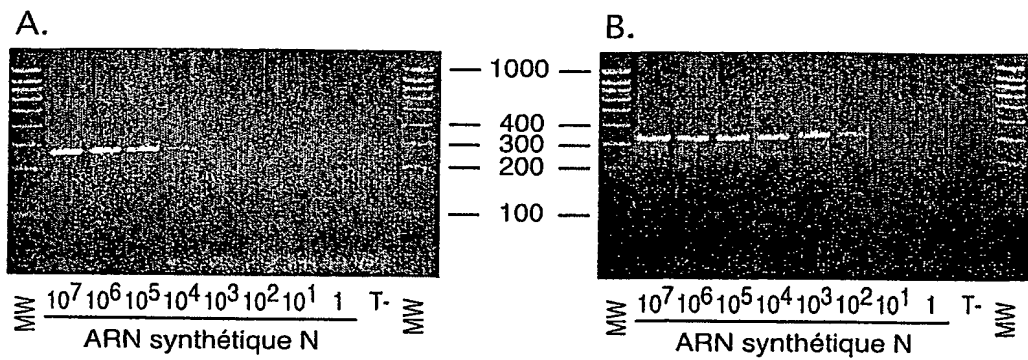


FIGURE 11

13/116

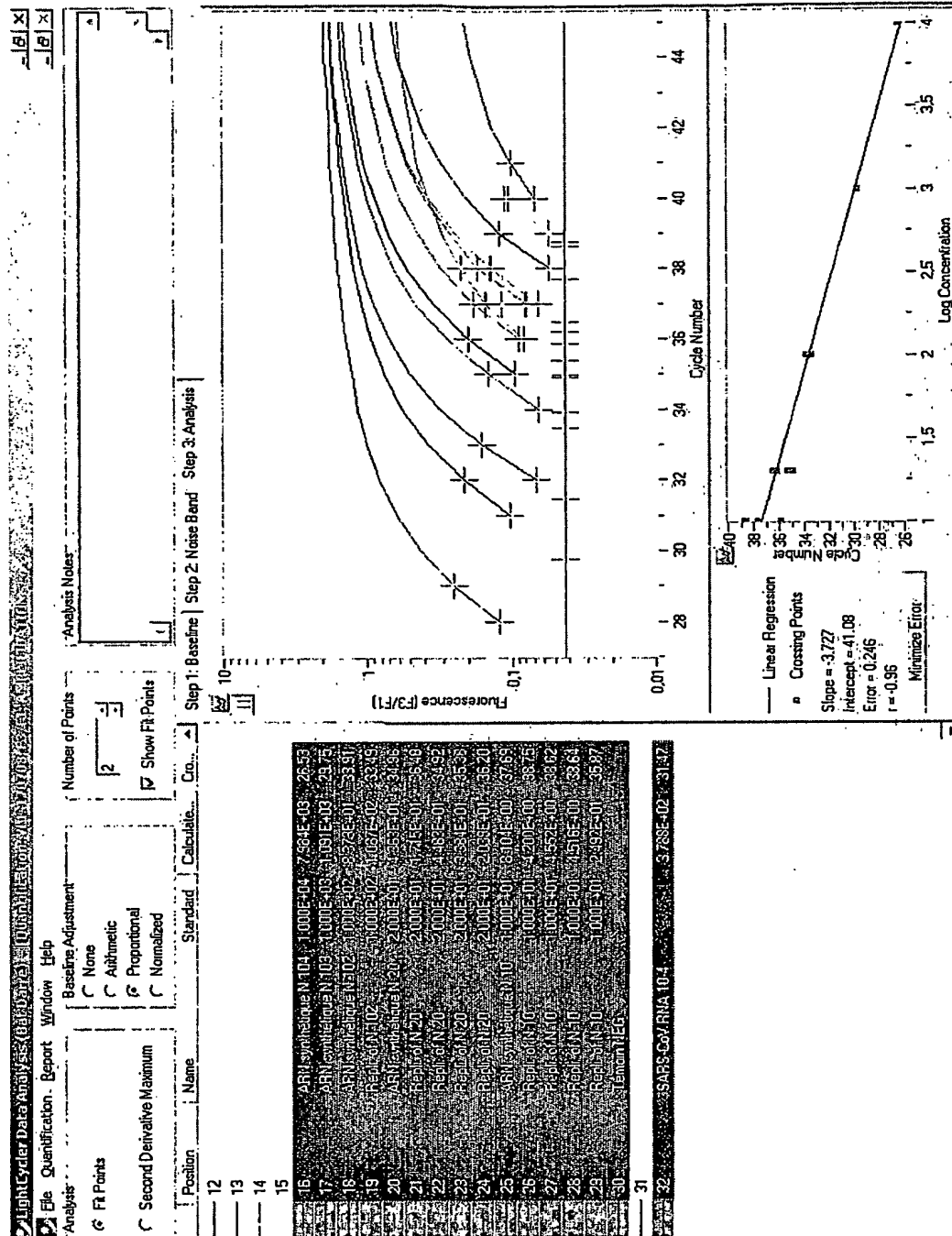


FIGURE 12

```

                >< ScrFI                >< XhoII
                >< MvaI                > < TthHB8I >< Sau3AI
                >< EcoRII                > < TaqI    >< NdeII
                >< Ecl136I                >< Sau3AI >< MflI
                >< DsaV                  >< NdeII  >< MboI
                >< BstOI                  >< MboI>< MnlI>< DpnII
                >< BstNI                  >< DpnII  >< BstYI
                >< BsiLI                  >< DpnI   >< BspAI
                >< BsaJI                  >< BspAI   >< Bsp143I
                >< ApyI                  >< Bsp143I>< BglII
ATATTAGGTT TTTACCTACC CAGGAAAAGC CAACCAACCT CGATCTCTTG TAGATCTGTT CTCTAAACGA
    10          20          30          40          50          60          70

                >< VneI
                >< SphI
                >< SnaI
                >< RmaI
                >< PaeI >< SduI
                >< NspI >< NspII
                >< NspHI >< HgiAI
                >< NlaIII >< Bsp1286I
                >< MaeI >< BmyI
                >< ApaLI
                >< MseI                >< BbvI                >< Alw44I
                >< DraI                >< AluI                > < Fnu4HI                >< Alw21I
ACTTTAAAAAT CTGTGTAGCT GTCGCTCGGC TGCATGCCTA GTGCACCTAC GCAGTATAAA CAATAATAAA
    80          90          100          110          120          130          140

                >< SfcI
                >< PstI
                >< MnlI
                >< Ksp632I
                >< HindII                > < MboII                >< EarI
                >< HincII                >< MaeIII                >< Eam1104I
TTTTACTGTC GTTGACAAGA AACGAGTAAC TCGTCCCTCT TCTGCAGACT GCTTACGGTT TCGTCCGTGT
    150          160          170          180          190          200          210

                >< TthHB8I                >< StyI
                >< TaqI                >< RmaI                >< ScrFI
                >< Sau3AI                >< MaeI                >< NciI
                >< NdeII                >< EcoT14I                >< MspI
                >< MboI                >< Eco130I                >< MaeIII
                >< DpnII                >< BssT1I                >< HpaII
                >< DpnI                >< BsaJI                >< HapII
                >< BspAI                >< BlnI                >< DsaV
                >< Bsp143I                >< AvrII                >< BcnI
TGCAGTCGAT CATCAGCATA CCTAGGTTTC GTCCGGGTGT GACCGAAAGG TAAGATGGAG AGCCTTGTTTC
    220          230          240          250          260          270          280

                >< RmaI
                >< HindII                >< MaeII> < Eco57I                >< Esp3I >< MaeII
                >< HincII                > < AflIII > < DdeI                >< BsmAI                >< MaeI
                >< Alw26I                >< BsmBI
TTGGTGTCAA CGAGAAAACA CACGTCCAAC TCAGTTTGCC TGTCCCTTCAG GTTAGAGACG TGCTAGTGCG
    290          300          310          320          330          340          350

```

FIGURE 13.1

15/116

```

    >< Sau96I
      >< PssI
    >< Pali
    >< NspIV
      >< MnlI
      >< HaeIII
    >< EcoO109I
      >< DraII>< MboII >< PmlI
    >< MnlI >< Cfr13I >< PmaCI
    >< Ksp632I >< BsuRI >< MaeII
    >< HinfI >< BsiZI>< EcoNI >< Eco72I
      >< EarI >< BshI >< BslI >< BsaAI
    >< PleI >< Eam1104I>< AsuI >< BsiYI>< BbrPI >< MnlI
TGGCTTCGGG GACTCTGTGG AAGAGGCCCT ATCGGAGGCA CGTGAACACC TCAAAAATGG CACTTGTGGT
  360          370          380          390          400          410          420

    >< RsaI
    >< Csp6I >< BspWI >< Tru9I
    >< MaeI >< AluI >< AfaI >< AluI >< SfaNI
CTAGTAGAGC TGGAAAAAGG CGTACTGCCC CAGCTTGAAC AGCCCTATGT GTTCATTAAA CGTTCTGATG
  430          440          450          460          470          480          490

    >< PalI
    >< HaeIII
    >< Tru9I >< GdiII >< RsaI
    >< MseI >< EaeI McrI ><
    >< Esp4I >< BsuRI >< BsmI BsiEI >< Csp6I
    >< AflII >< BshI >< AluI >< BscCI >< AfaI
CCTTAAGCAC CAATCACGGC CACAAGGTCG TTGAGCTGGT TGCAGAAATG GACGGCATTG AGTACGGTCG
  500          510          520          530          540          550          560

    >< NspI
    >< ScaI >< NspHI
    >< RsaI >< NlaIII
    >< Csp6I >< BslI >< MboII
    >< BsrI >< BsiYI >< MboII
    >< AciI >< AfaI >< AflIII >< MunI >< AciI
TAGCGGTATA AACTGAGGAG TACTCGTGCC ACATGTGGGC GAAACCCCAA TTGCATACCG CAATGTTCTT
  570          580          590          600          610          620          630

    >< TthHB8I
    >< TaqI
    >< Sau3AI
    >< NdeII
    >< MboI
    >< DpnII
    >< DpnI
    >< ClaI
    >< Bsu15I
    >< BspDI
    >< BspAI
    >< Bsp143I
    >< Bsp106I
    >< BsiXI MaeIII >
    >< BscI>< SfaNI DdeI ><
    >< BscBI >< AluI >< BanIII BfrI ><
CTTCGTAAGA ACGGTAATAA GGGAGCCGGT GGTCATAGCT ATGGCATCGA TCTAAAGTCT TATGACTTAG
  640          650          660          670          680          690          700

```

FIGURE 13.2

16/116

```

    >< Sau3AI
    >< NdeII
    >< MboI
    >< HphI
    >< DpnII
    >< BspAI
    >< AlwI>< DpnI
    >< AluI
    >< Bsp143I
    >< MboII
    >< BsrI
    >< DdeI
    >< VneI ><
    >< SnoI ><
    >< NlaIII
    >< ApaLI ><
    >< Alw44I ><
    GTGACGAGCT TGGCACTGAT CCCATTGAAG ATTATGAACA AACTGGAAC ACTAAGCATG GCAGTGGTGC
    710          720          730          740          750          760          770

    >< SstI
    >< SduI
    >< SacI
    >< NspII
    >< MnlI
    >< HgiAI
    >< Eco24I
    >< TaqI
    >< TthHB8I
    >< Pali ><
    >< NspII
    >< Ecl136II
    >< SalI
    >< NspIV ><
    >< HgiAI
    >< Bsp1286I
    >< RtrI
    >< HaeIII ><
    >< Cfr13I ><
    >< DraIII
    >< BmyI
    >< HindII
    >< BsuRI ><
    >< Bsp1286I
    >< BanII
    >< HincII
    >< BsiZI ><
    >< BmyI
    >< Alw21I
    >< BsgI
    >< BshI ><
    >< Alw21I
    >< AluI
    >< MaeIII
    >< AccI
    >< AsuI ><
    ACTCCGTGAA CTCACCTCGTG AGCTCAATGG AGGTGCAGTC ACTCGCTATG TCGACAACAA TTTCTGTGGC
    780          790          800          810          820          830          840

    >< ThaI
    >< ThaI
    >< MvnI
    >< MvnI
    >< HinPII
    >< Hin6I
    >< VneI
    >< HhaI
    >< SnoI
    >< Eco64I
    >< CfoI
    >< SduI
    >< Csp6I
    >< BstUI
    >< NspII ><
    >< BscBI
    >< BstUI
    >< HgiAI ><
    >< BanI
    >< Bsp50I
    >< Bsp1286I ><
    >< Asp718
    >< Bsp50I
    >< BmyI
    >< AfaI
    >< AciI
    >< ApaLI
    >< AccB1I
    >< AccII
    >< Alw44I
    >< Acc65I
    >< MnlI
    >< SfaNI
    >< AccII
    >< Alw21I ><
    CCAGATGGGT ACCCTCTTGA TTGCATCAAA GATTTTCTCG CACGCGCGGG CAAGTCAATG TGCACTCTTT
    850          860          870          880          890          900          910

    >< TthHB8I
    >< TthHB8I
    >< TaqI
    >< TaqI
    >< MnlI
    >< Ksp632I
    >< NlaIII ><
    >< HinfI>< PleI
    >< NlaIII
    >< Eam1104I
    >< MboII
    >< MaeIII
    >< EcoRII ><
    >< EarI
    >< BbvI>< AccI >< Fnu4HI
    >< DsaV ><
    CCGAACAACT TGATTACATC GAGTCGAAGA GAGGTGTCTA CTGCTGCCGT GACCATGAGC ATGAAATTGC
    920          930          940          950          960          970          980

    >< TthHB8I
    >< TaqI
    >< SfuI
    >< NspV>< Tru9I
    >< LspI>< MseI
    >< ScrFI
    >< HinPII

```

FIGURE 13.3

17/116

```

>< MvaI      >< Hin6I      >< SduI      >< Csp45I
>< Ecl136I   >< HhaI      >< NspII     >< BstBI
>< BstOI     >< HaeII     >< HgiAI     >< Bsp119I
>< BstNI     >< Eco47III   >< Bsp1286I  >< BsiCI
>< BsiLI     >< CfoI      >< BmyI      >< Bpu14I
>< ApyI >< DdeI >< Bsp143II >< AluI   >< Alw21I   >< AsuII
CTGGTTCAC T GAGCGCTCTG ATAAGAGCTA CGAGCACCAG ACACCCTTCG AAATTAAGAG TGCCAAGAAA
      990      1000      1010      1020      1030      1040      1050

                                >< Tru9I
                                >< MseI
                                >< BsmI
                                >< BscCI
                                >< MnlI
TTTGACACTT TCAAAGGGGA ATGCCCAAAG TTTGTGTTTC CTCTTAACTC AAAAGTCAAA GTCATTCAAC
      1060      1070      1080      1090      1100      1110      1120

>< PmlI
>< PmaCI
>< MaeII
>< Eco72I
>< BsaAI
>< BbrPI
>< AflIII   >< MnlI>< DdeI   >< NlaIII   >< RsaI
>< AccI     >< Bst1107I >< Csp6I
CACGTGTTGA AAAGAAAAAG ACTGAGGGTT TCATGGGGCG TATACGCTCT GTGTACCCTG TTGCATCTCC
      1130      1140      1150      1160      1170      1180      1190

>< SfaNI
>< MaeIII   >< AccI
ACAGGAGTGT AACAAATATGC ACTTGTCTAC CTTGATGAAA TGTAATCATT GCGATGAAGT TTCATGGCAG
      1200      1210      1220      1230      1240      1250      1260

                                >< SinI
                                >< Sau96I
                                PssI ><
                                >< Psp5II
                                >< PpuMI
                                >< NspIV
                                >< NspHII
                                >< Eco47I
                                >< DraII
                                >< Cfr13I
                                >< BsiZI
                                >< Bme18I
                                >< AvaII
                                >< AsuI
>< MaeII
ACGTGCGACT TTCTGAAAGC CACTTGTGAA CATTGTGGCA CTGAAAATTT AGTTATTGAA GGACCTACTA
      1270      1280      1290      1300      1310      1320      1330

                                Eco0109I >< AflIII >
                                Van91I ><
                                SinI ><
                                Sau96I ><
                                PflMI ><
                                NspIV ><
                                NspHII >
                                Eco47I ><
                                Cfr13I ><
                                BslI ><
                                BsiZI ><
                                BsiYI ><
                                Bme18I ><
                                AvaII ><
                                AsuI ><

>< RsaI
>< NspI
>< NlaIV
>< NlaIII
>< NspHI>< KpnI
>< Eco64I
>< Csp6I
>< BscBI
>< BanI
>< Asp718
>< AfaI
>< AccBI

```

FIGURE 13. 4


```

    >< Acc65I          >< SfcI          >< NlaIII          AccB7I ><
CATGTGGGTA CCTACCTACT AATGCTGTAG TGAAAATGCC ATGTCCTGCC TGTCAAGACC CAGAGATTGG
    1340          1350          1360          1370          1380          1390          1400

                                >< TthHB8I
                                >< TaqI>< MnlI
                                >< HinfI

    >< DdeI          >< PleI          >< AciI
ACCTGAGCAT AGTGTTCAG ATTATCACAA CCACTCAAAC ATTGAACTC GACTCCGCAA GGGAGGTAGG
    1410          1420          1430          1440          1450          1460          1470

    >< RmaI          NlaIV ><
    >< MnlI          >< BsrI
    >< MaeI          >< BbvI          >< Fnu4HI          BscBI ><
ACTAGATGTT TTGGAGGCTG TGTGTTTGCC TATGTTGGCT GCTATAATAA GCGTGCCTAC TGGGTTCCCTC
    1480          1490          1500          1510          1520          1530          1540

                                XhoII ><
                                Sau3AI ><
                                NdeII ><
                                MflI ><
                                MboI ><
                                >< MaeIII          >< Eco31I          DpnII ><
                                >< Pali          >< HaeIII          >< BsrI          >< MnlI DpnI >
                                >< BsuRI          >< BsmAI          BstYI ><
    >< RmaI          >< DdeI          >< BspWI          >< BsaI>< HphI          BspAI ><
    >< MnlI          >< BshI>< BglI          >< Alw26I          Bsp143I >
    >< MaeI          GTGCTAGTGC TGATATTGGC TCAGGCCATA CTGGCATTAC TGGTGACAAT GTGGAGACCT TGAATGAGGA
    1550          1560          1570          1580          1590          1600          1610

                                > < Tru9I
                                > < MseI
                                >< MaeII          >< Tru9I
                                >< HpaI          > < MnlI
                                >< HindII          > < Ksp632I
                                >< HinfI >< PleI >< HincII          > < EarI
                                >< AlwI          >< DdeI          >< AflIII          >< MseI          > < Eam1104I
    TCTCCTTGAG ATACTGAGTC GTGAACGTGT TAACATTAAC ATTGTTGGCG ATTTTCATTT GAATGAAGAG
    1620          1630          1640          1650          1660          1670          1680

                                >< MboII          PleI ><
                                >< BstXI          >< SfaNI          > < HinfI
    GTTGCCATCA TTTTGGCATC TTTCTCTGCT TCTACAAGTG CCTTTATTGA CACTATAAAG AGTCTTGATT
    1690          1700          1710          1720          1730          1740          1750

                                >< StyI
                                >< MaeIII
                                >< EcoT14I
                                >< PleI          >< Eco130I
                                >< MaeIII          >< BssT1I          BslI ><
                                >< HinfI>< AciI          >< BsaJI          BsiYI ><
    ACAAGTCTTT CAAAACCATT GTTGAGTCCT GCGGTAAC TAAGTTACC AAGGGAAAGC CCGTAAAAGG
    1760          1770          1780          1790          1800          1810          1820

                                >< Sau3AI          >< Van91I
                                >< NdeII          >< PflMI
                                >< MboI          >< DraIII
                                >< DpnII          >< BslI
                                >< DpnI >< Tru9I          >< BsiYI
                                >< BspAI >< MseI          >< BbvI          >< MnlI
                                >< Bsp143I          >< AccB7I          Fnu4HI ><

```

FIGURE 135

19/116

```

TGCTTGAAC ATTGGACAAC AGAGATCAGT TTTAACACCA CTGTGTGGTT TTCCCTCACA GGCTGCTGGT
  1830      1840      1850      1860      1870      1880      1890

      >< ThaI
      >< SfaNI
      >< MvnI
      >< HinPII
    >< HinPII
    >< Hin6I
    >< Hin6I
      >< HhaI
    >< Sau3AI      >< HhaI
    >< NdeII      >< CfoI
    >< MboI      >< CfoI
    >< DpnII      >< BstUI
      >< DpnI    >< BssHII
    >< BspAI      >< Bsp50I
      >< Bsp143I  >< AccII
    >< Fnu4HI      >< BbvI
GTTATCAGAT CAATTTTTCG GCGCACACTT GATGCAGCAA ACCACTCAAT TCCTGATTTG CAAAGAGCAG
  1900      1910      1920      1930      1940      1950      1960

      >< TthHB8I
      >< StyI
      >< NcoI
    >< HindII
    >< HincII
      >< HinfI
      >< EcoT14I
      >< Eco57I
    >< TaqI>< Eco130I
    >< SalI >< DsaI
    >< RtrI >< BssT1I
      >< BsaHI
      >< BbiII>< NlaIII
      >< AcyI >< HgaI
    >< MaeIII
      >< BbvI
    >< MaeII >< AccI>< BsaJI      HphI ><
CTGTCACCAT ACTTGATGGT ATTTCTGAAC AGTCATTACG TCTTGTGCGAC GCCATGGTTT ATACTTCAGA
  1970      1980      1990      2000      2010      2020      2030

      >< RsaI
      >< Csp6I
    >< BspMI      >< MaeIII >< BsrI >< AfaI      >< DdeI
CCTGCTCACC AACAGTGTC TATTATGGC ATATGTAACT GGTGGTCTTG TACAACAGAC TTCTCAGTGG
  2040      2050      2060      2070      2080      2090      2100

      >< StuI
      >< PstI
      >< HaeIII
      >< Eco147I
    >< SduI      >< DdeI
    >< NspII      >< BsuRI
    >< Bsp1286I    >< BshI
    >< BmyI      >< AatI      >< MnlI      DdeI ><
    >< BfrI ><
TTGTCTAATC TTTTGGGCAC TACTGTTGAA AAACCTCAGGC CTATCTTTGA ATGGATTGAG GCGAAACTTA
  2110      2120      2130      2140      2150      2160      2170

      >< TfiI
      >< HinfI
      >< FokI
    >< SfaNI >< BsgI
    >< Tth111I ><
    >< AspI ><
GTGCAGGAGT TGAATTTCTC AAGGATGCTT GGGAGATTCT CAAATTTCTC ATTACAGGTG TTTTGTGACAT
  2180      2190      2200      2210      2220      2230      2240

```

FIGURE 13.6

Tru9I ><
MseI ><
HpaI >
HindII >
HincII >

```
>< Sau3AI
>< NdeII
>< MboI
    > < MaeIII
>< FbaI
>< DpnII
>< DpnI
    >< Sau3AI
    >< NdeII
    >< DpnII
    >< DpnIMboII ><
```

>< HinP1I DdeI ><
>< HincII >< Bsp143I

```

>< MaeII          >< PvuII
>< Bst1107I       >< Psp5I
>< BsaAI          >< NspBII
>< BbvI           Fnu4HI ><
>< BbvI           > < Fnu4HI

```

```
>> AccI          >> AluI
TATACGT GGCAAGGAGC AGCTGCAACT
      2430          2440          2450
```

```

> < XhoI
>< TthHB8I
>< TthHB8I>< TaqI
> < SfaI
> < PaeR7I
> < NspIII
>< HphI >< HinII
> < Eco88I
> < CcrI
>< Esp3I >< BsaHI
> < BcoI
>< BsmAI >< BbiII
> < AvaI >> HgaI
>< TaqI > < Ama87I>< BsmBI

```

FIGURE 13.7

21/116

```

                >< Pali >< NlaIII
                >< HaeIII >< MnlI
                >< BsuRI >< DdeI >< Tru9I
                >< BshI >< BfrI >< MseI
>< AluI          >< BsrI          >< BshI >< BfrI >< MseI
GAGCTATCGT TGGCACACCA GTCTGTGTAA ATGGCCTCAT GCTCTTAGAG ATTAAGGACA AAGAACAATA
    2600          2610          2620          2630          2640          2650          2660

                >< VneI
                Tru9I ><
                >< SnoI
                >< SduI
                >< NspII
                MseI ><
                >< HgiAI
                Bsp1286I >< BslI ><
                BsiYI ><
                >< BmyI
                >< ApaLI
                >< Tru9I >< Alw44I
                >< MseI >< Alw21I
CTGCGCATTG TCTCCTGGTT TACTGGCTAC AAACAATGTC TTTCGCTTAA AAGGGGGTGC ACCAATTAAA
    2670          2680          2690          2700          2710          2720          2730

                >< TfiI
>< MaeIII          >< MboII > < MaeIII >< HinfI AluI ><
GGTGTAACCT TTGGAGAAGA TACTGTTTGG GAAGTTCAAG GTTACAAGAA TGTGAGAATC ACATTTGAGC
    2740          2750          2760          2770          2780          2790          2800

                >< RsaI
                >< NlaIV
                MaeIII ><
                >< MspI >< KpnI
                >< HpaII
                >< HapII
                > < Eco64I
                >< SduI >< Csp6I
                >< NspII >< TfiI >< BscBI
                >< HgiAI > < BanI
                >< Bsp1286I > < Asp718
                >< BmyI >< HinfI >< AfaI
                >< Alw21I > < AccB1I
                >< AccI > < Acc65I
TTGATGAACG TGTTGACAAA GTGCTTAATG AAAAGTGCTC TGTCTACACT GTTGAATCCG GTACCGAAGT
    2810          2820          2830          2840          2850          2860          2870

                >< Sau3AI
                >< NdeII
                >< MboI
                >< DpnII
                > < DpnI
                >< NspI
                >< NspHI
                >< NlaIII
                >< MboII >< BspAI
                > < BsrI > < Bsp143I
>< DdeI          >< MnlI >< AlwNI >< BbsI >< AlwNI
TACTGAGTTT GCATGTGTTG TAGCAGAGGC TGTGTGAAG ACTTTACAAC CAGTTTCTGA TCTCCTTACC
    2880          2890          2900          2910          2920          2930          2940

                >< Sau3AI
                >< NdeII
                >< MboI
                >< DpnII
                >< DpnI
                >< BspAI

```

FIGURE 13.8

22/11 6

```

    >< NlaIII>< BspI43I          >< AluI          >< SfaNI
AACATGGGTA TTGATCTTGA TGAGTGGAGT GTAGCTACAT TCTACTTATT TGATGATGCT GGTGAAGAAA
    2950          2960          2970          2980          2990          3000          3010

                                >< SfaNI
                                >< MnlI
                                >< Ksp632I          >< MnlI
                                >< EarI          >< MboII
    >< MboII          >< GsuI          >< MnlI          >< MboII
                                >< Eam1104I          >< MboII
    >< HphI          >< MaeII>< BpmI          >< MnlI          >< MboII
ACTTTTCATC ACGTATGTAT TGTTCTTTT ACCCTCCAGA TGAGGAAGAA GAGGACGATG CAGAGTGTGA
    3020          3030          3040          3050          3060          3070          3080

                                >< RsaI
                                >< RsaI
    >< NlaIII
                                >< MnlI          >< FokI
                                >< Csp6I          Eco31I ><
                                >< Csp6I          >< MamI          BsmAI ><
                                >< MboII          >< AfaI          >< BsiBI          BsaI ><
                                >< MboII          >< AfaI          >< BsaBIAIw26I ><
GGAAGAAGAA ATTGATGAAA CCTGTGAACA TGAGTACGGT ACAGAGGATG ATTATCAAGG TCTCCCTCTG
    3090          3100          3110          3120          3130          3140          3150

    >< NlaIV>< PvuII>< XmnI
    >< Eco64I >< Psp5I          >< TthHB8I
    >< MnlI >< DdeI          >< TaqI          >< MnlI          >< MboII
    >< BscBI>< NspBII >< MnlI          >< Ksp632I          >< MboII >< MboII
    >< BanI          >< MnlI          >< EarI          >< BsrI
    >< AccB1I >< AluI >< Asp700I          >< Eam1104I >< MboII>< BbsI
GAATTTGGTG CCTCAGCTGA AACAGTTCGA GTTGAGGAAG AAGAAGAGGA AGACTGGCTG GATGATACTA
    3160          3170          3180          3190          3200          3210          3220

                                >< Tru9I
    >< FokI          >< MseI          >< Eco57I
    >< DdeI          >< BsrI>< MboII          BsrI ><
CTGAGCAATC AGAGATTGAG CCAGAACCAG AACCTACACC TGAAGAACCA GTTAATCAGT TTAATGGTTA
    3230          3240          3250          3260          3270          3280          3290

    >< Tru9I          >< MnlI
    >< MseI          >< Tru9I          >< HindII>< Tru9I          >< DraIII
    >< DraI          >< MseI          >< HincII>< MseI          >< BspWI
TTTAAACTT ACTGACAATG TTGCCATTAA ATGTGTTGAC ATCGTTAAGG AGGCACAAAG TGCTAATCCT
    3300          3310          3320          3330          3340          3350          3360

                                >< VneI
                                >< SnoI
                                >< SduI
                                >< NspII
                                >< HgiAI
                                >< Bsp1286I
                                >< BmyI
                                >< ApaLI
                                >< Alw44I
    >< HphI          >< NlaIII          >< Alw21I
    >< BbvI          >< Fnu4HI          >< BspMI
ATGGTGATTG TAAATGCTGC TAACATACAC CTGAAACATG GTGGTGGTGT AGCAGGTGCA CTCAACAAGG
    3370          3380          3390          3400          3410          3420          3430

                                >< Sau96I
                                >< Pali
                                >< NspIV
                                >< HaeIII
                                >< Cfr13I
    >< NlaIV

```

FIGURE 13.9

23/116

```

>< Eco64I
>< BscBI
>< BanI
>< AccBII>< NlaIII
CAACCAATGG TGCCATGCAA AAGGAGAGTG ATGATTACAT TAAGCTAAAT GGCCCTCTTA CAGTAGGAGG
3440 3450 3460 3470 3480 3490 3500

>< BsuRI
>< BsiZI
>< BshI
>< MnlI
>< AluI
>< AsuI
>< MnlI
>< SinI
>< Sau96I
>< NspIV
>< NspHI>< NspHII
>< Eco47I
>< Cfr13I
>< NlaIII >< BspMI
>< BsiZI
>< Bmel8I
>< AvaII MnlI ><
>< DdeI
>< NspI>< AsuI FokI ><
GTCTTGTTTG CTTTCTGGAC ATAATCTTGC TAAGAAGTGT CTGCATGTTG TTGGACCTAA CCTAAATGCA
3510 3520 3530 3540 3550 3560 3570

>< Tru9I
>< HphI> < MseI
>< Esp4I
>< AluI >< NdeI
>< AflIII>< Fnu4HI >< BbvI
GGTGAGGACA TCCAGCTTCT TAAGGCAGCA TATGAAAATT TCAATTCACA GGACATCTTA CTTGCACCAT
3580 3590 3600 3610 3620 3630 3640

RsaI ><
Csp6I ><
AfaI ><
>< Eco57I
>< BcgI
TGTTGTCAGC AGGCATATTT GGTGCTAAAC CACTTCAGTC TTTACAAGTG TGCCTGCAGA CGGTTCGTAC
3650 3660 3670 3680 3690 3700 3710

>< BsgI
>< BcgI/a
>< BspMI
>< AluI
>< NlaIII
ACAGGTTTAT ATTGCAGTCA ATGACAAAGC TCTTTATGAG CAGGTTGTCA TGGATTATCT TGATAACCTG
3720 3730 3740 3750 3760 3770 3780

>< MnlI
>< RmaI >< MnlI >< NlaIV >< TfiI >< MboII
>< MaeI >< Eco57I >< BscBI >< HinfI >< DdeI
AAGCCTAGAG TGGAAGCACC TAAACAAGAG GAGCCACCAA ACACAGAAGA TTCCAAAACCT GAGGAGAAAT
3790 3800 3810 3820 3830 3840 3850

>< Tru9I
>< StuI
>< Pali
>< MseI >< MnlI >< MaeIII
>< HaeIII >< EcoO65I
>< Eco147I >< Eco91I
>< BsuRI BstXI ><
>< Csp6I >< TthHB8I >< BshI >< BstPI
>< AfaI >< TaqI >< AatI >< BstEII
CTGTCGTACA GAAGCCTGTC GATGTGAAGC CAAAAATTAA GGCCTGCATT GATGAGGTTA CCACAACACT
3860 3870 3880 3890 3900 3910 3920

TfiI ><
NlaIII ><
HinfI ><
>< DdeI
>< EcoRV >< HindIII

```

FIGURE 13.10

```

>< BsrI      >< MboII      >< MaeIII      >< Eco32I      >< AluI
GGAAGAACT AAGTTTCTTA CCAATAAGTT ACTCTTGTTT GCTGATATCA ATGGTAAGCT TTACCATGAT
  3930      3940      3950      3960      3970      3980      3990

      >< NspI
      >< NspHI
      >< NlaIII
    >< MnlI      >< SfaNI
      >< DdeI      >< EcoNI
      >< BfrI      >< MboII >< BslI      >< NlaIII
>< DdeI      >< HphI      >< BsiYI      >< FokI
TCTCAGAACA TGCTTAGAGG TGAAGATATG TCTTTCCTTG AGAAGGATGC ACCTTACATG GTAGGTGATG
  4000      4010      4020      4030      4040      4050      4060

    >< SpeI
    >< RmaI
    >< MaeI      >< EcoRV>< HphI      >< SfaNI
    >< HphI      >< Eco32I      >< MnlI      >< DdeI
TTATCACTAG TGGTGATATC ACTTGTGTTG TAATACCCTC CAAAAAGGCT GGTGGCACTA CTGAGATGCT
  4070      4080      4090      4100      4110      4120      4130

      >< ScrFI
      >< RsaI
      >< MvaI
      >< EcoRII
      >< Ecl136I
      >< DsaV
    >< Csp6I >< EcoNI
      >< BstOI
      >< BstNI
      >< BsiLI
      >< BsaJI
      >< BsaAI >< BslI
      >< MaeII>< ApyI
      >< AfaI >< BsiYI
    >< AluI      >< MboII      >< BsrI
CTCAAGAGCT TTGAAGAAAG TGCCAGTTGA TGAGTATATA ACCACGTACC CTGGACAAGG ATGTGCTGGT
  4140      4150      4160      4170      4180      4190      4200

      >< Tru9I
      >< MseI
      >< DdeI      >< Esp4I
    >< MnlI      >< BspWI
    >< FokI      >< AluI      >< AflII      >< Eco57I >< AfaI
TATACACTTG AGGAAGCTAA GACTGCTCTT AAGAAATGCA AATCTGCATT TTATGTACTA CCTTCAGAAG
  4210      4220      4230      4240      4250      4260      4270

      >< ScrFI
      >< MvaI
      >< EcoRII
      >< Ecl136I
      >< XmnI
    >< Ksp632I >< RmaI      >< DsaV      NlaIII ><
    >< EarI >< TfiI>< MboII      >< BstOI      Ksp632I ><
    >< Eam1104I >< MaeI      >< BstNI      EarI
    >< DdeI >< HinfI      >< BsiLI      Eam1104I ><
    >< BspWI      >< Asp700I      >< ApyI      BsmAI ><
    >< Alw26I ><
CACCTAATGC TAAGGAAGAG ATTCTAGGAA CTGTATCCTG GAATTTGAGA GAAATGCTTG CTCATGCTGA
  4280      4290      4300      4310      4320      4330      4340

      >< VspI      >< Zsp2I
      >< Tru9I      >< Ppu10I
      >< MseI      >< NsiI
    >< MboII      >< NlaIII      >< FokI
      >< Eco57I >< Mph1103I >< FokI

```

FIGURE 13. 11

```

                >< AsnI          >< EcoT22I          >< BspWI
                >< AseI          >< AvaIII          >< BglI          >< MaeII
AGAGACAAGA AAATTAATGC CTATATGCAT GGATGTTAGA GCCATAATGG CAACCATCCA ACGTAAGTAT
    4350      4360      4370      4380      4390      4400      4410

                >< SfaNI
                > < HindII          >< TfiI          >< SpeI
                > < HincII>< MboII          >< RmaI
                >< MnlI          >< DrdI >< HinfI          >< MaeI
AAAGGAATTA AAATTCAAGA GGGCATCGTT GACTATGGTG TCCGATTCTT CTTTATACT AGTAAAGAGC
    4420      4430      4440      4450      4460      4470      4480

                >< MaeIII
>< SfcI          >< Fnu4HI          >< MunI
                >< AluI          >< AluI          >< AciI          MaeIII ><
CTGTAGCTTC TATTATTACG AAGCTGAACT CTCTAAATGA GCCGCTTGTC ACAATGCCAA TTGGTTATGT
    4490      4500      4510      4520      4530      4540      4550

                >< ThaI
                >< MvnI
                >< MboII
                >< HinfII
>< HinfII
                >< HinfI
                >< HinfI
                >< HhaI
                >< HhaI
                >< Fnu4HI
                >< CfoI
                >< CfoI
                >< BstUI
                >< BssHII>< BspWI          >< Tru9I
                >< Bsp50I          >< MseI
                >< BbvI          >< AccII          >< AluI          HphI ><
GACACATGGT TTTAATCTTG AAGAGGCTGC GCGCTGTATG CGTTCTCTTA AAGCTCCTGC CGTAGTGTCA
    4560      4570      4580      4590      4600      4610      4620

                >< MaeIII
>< SfaNI          >< AlwNI          >< MnlI >< MnlI>< DdeI
GTATCATCAC CAGATGCTGT TACTACATAT AATGGATACC TCACTTCGTC ATCAAAGACA TCTGAGGAGC
    4630      4640      4650      4660      4670      4680      4690

                >< SinI
                >< Sau96I
                >< NspIV
                >< NspHII
>< SduI          >< Eco47I
>< NspII          >< Cfr13I
>< HgiAI          >< Bsi2I
>< Bsp1286I          >< Bme18I          >< RsaI
>< BmyI          >< AvaII          >< Csp6I
>< Alw21I          >< AsuI          >< AfaI
ACTTTGTTAGA AACAGTTTCT TTGGCTGGCT CTTACAGAGA TTGGTCCTAT TCAGGACAGC GTACAGAGTT
    4700      4710      4720      4730      4740      4750      4760

                > < TthHB8I
                > < TaqI
                >< SduI
                >< NspII
                >< Van91I
                >< RsaI          >< PflMI          >< Eco24I
                >< HphI          >< BslI          >< Bsp1286I
                >< Csp6I          >< BsiYI          >< BmyI          >< GsuI ><
                >< Tru9I
                >< MseI
                >< Esp4I

```

FIGURE 13.12

26/116

```

                >< AflIII >< MaeIII                >< AfaI >< AccB7I >< BanIIBpmI ><
AGGTGTTGAA TTTCTTAAGC GTGGTGACAA AATTGTGTAC CACACTCTGG AGAGCCCCGT CGAGTTTCAT
    4770          4780          4790          4800          4810          4820          4830

                                >< Tru9I
                                >< PleI >< EcoNI
                                >< MnlI >< BslI
                                >< BsmAI >< BsiYI
>< MnlI                >< HphI                >< HinfI>< Alw26I>< AciI >< MseI
CTTGACGGTG AGGTTCTTTC ACTTGACAAA CTAAAGAGTC TCTTATCCCT GCGGGAGGTT AAGACTATAA
    4840          4850          4860          4870          4880          4890          4900

                                >< AluI                >< NdeI
AAGTGTTTAC AACTGTGGAC AACACTAATC TCCACACACA GCTTGTGGAT ATGTCTATGA CATATGGACA
    4910          4920          4930          4940          4950          4960          4970

    >< SinI
    >< Sau96I
    >< NspIV
    >< NspHII
    >< Eco47I
    >< Cfr13I
    >< BsiZI
    >< Bmel8I
    >< AvaII
    >< AsuI
                                >< MaeIII >< Tru9I >< MnlI
                                >< FokI >< MseI >< BspHI
GCAGTTTGGT CCAACATACT TGGATGGTGC TGATGTTACA AAAATTAAAC CTCATGTAAA TCATGAGGGT
    4980          4990          5000          5010          5020          5030          5040

                                > < TthHB8I
                                > < TaqI
    >< RsaI
    > < RmaI                >< SnaBI.                >< ScaI
    > < MaeI                >< MaeII >< HindIII >< RsaI
    >< Csp6I                >< Eco105I                >< Csp6I
    >< AfaI                >< BsaAI >< AluI >< AfaI
AAGACTTTCT TTGTACTACC TAGTGATGAC ACACTACGTA GTGAAGCTTT CGAGTACTAC CATACTCTTG
    5050          5060          5070          5080          5090          5100          5110

    >< RsaI
    >< NspI
    >< NspHI
    >< NlaIII
    > < Csp6I                >< Tru9I                MnlI >
    >< AflIII                >< MseI                BslI ><
    >< AfaI                >< DraI                BsiYI ><
ATGAGAGTTT TCTTGGTAGG TACATGTCTG CTTTAAACCA CACAAAGAAA TGGAAATTTC CTCAAGTTGG
    5120          5130          5140          5150          5160          5170          5180

    >< Tru9I >< Tru9I                >< RmaI
    >< MseI >< MseI                >< MunI                >< MaeI                AluI >
TGGTTTAACT TCAATTAAAT GGGCTGATAA CAATTGTTAT TTGTCTAGTG TTTTATTAGC ACTTCAACAG
    5190          5200          5210          5220          5230          5240          5250

                                >< SfaNI
                                >< SduI
                                >< NspII
                                >< Eco24I
                                >< Bsp1286I
                                >< BmyI                HphI >
                                >< BbvI Fnu4HI ><
                                >< BanII >< BspWI
    >< MnlI

```

FIGURE 13.13

27/116

```

CTTGAAGTCA AATTCAATGC ACCAGCACTT CAAGAGGCTT ATTATAGAGC CCGTGCTGGT GATGCTGCTA
5260          5270          5280          5290          5300          5310          5320

>< VneI
>< SnoI
    >< SduI
    >< NspII
    >< HgiAI
    >< BspI286I
    >< BmyI
>< ApaLI
>< Alw44I
    >< Alw21I
    >< AluI
    MboII ><
    >< HphI
ACTTTTGTGC ACTCATACTC GCTTACAGTA ATAAAACTGT TGGCGAGCTT GGTGATGTCA GAGAAACTAT
5330          5340          5350          5360          5370          5380          5390

    > < SphI
    > < PaeI
    > < NspI
    > < NspHI >< TfiI
    >< Tru9I
    >< SfcI > < NlaIII>< HinfI
    >< MseI
GACCCATCTT CTACAGCATG CTAATTTGGA ATCTGCAAAG CGAGTTCTTA ATGTGGTGTG TAAACATTGT
5400          5410          5420          5430          5440          5450          5460

    >< RsaI
    >< Tru9I
    > < Csp6I
    Esp4I >
    >< MseI
    >< AluI
    >< AfaI
    AflII >
GGTCAGAAAA CTACTACCTT AACGGGTGTA GAAGCTGTGA TGTATATGGG TACTCTATCT TATGATAATC
5470          5480          5490          5500          5510          5520          5530

    >< RsaI
    >< MboII
    >< RmaI HinfI ><
    >< Csp6I
    >< MaeI >< BbsI
    >< AfaI
>< Tru9I
    >< SfaNI
    >< NlaIII
    >< AfaI
TTAAGACAGG TGTTTCCATT CCATGTGTGT GTGGTCGTGA TGCTACACAA TATCTAGTAC AACAAGAGTC
5540          5550          5560          5570          5580          5590          5600

    >< RsaI
    >< Csp6I
    >< AfaI
    >< PleI
    > < DdeI
    >< BspWI >< BspMI
    >< AfaI
TTCTTTTGTG ATGATGTCTG CACCACCTGC TGAGTATAAA TTACAGCAAG GTACATTCTT ATGTGCGAAT
5610          5620          5630          5640          5650          5660          5670

    >< Eco31I
    >< DdeI
    >< BsmAI
    >< BsaI
    MnlI ><
    >< AfaI >< BsrI
    >< Alw26I
    HphI >
GAGTACACTG GTAACATCA GTGTGGTCAT TACTCTCATA TAACTGCTAA GGAGACCCTC TATCGTATTG
5680          5690          5700          5710          5720          5730          5740

    >< SstI
    >< SnuI
    >< SacI
    >< NspII
    >< HgiAI
    >< Eco24I
    >< Ecl136II
    >< BspI286I
    >< BmyI

    >< SinI
    >< Sau96I
    >< NspIV
    >< NspHII
    > < RsaI >< MaeIII
    >< Eco47I
    >< Cfr13I
    >< BsiZI
    >< BmeI8I

```

FIGURE 13. 14

28/116

```

    >< BanII
    >< Alw21I
    >< AluI
ACGGAGCTCA CCTTACAAAG ATGTCAGAGT ACAAAGGACC AGTGAAGTAT GTTTTCTACA AGGAAACATC
    5750      5760      5770      5780      5790      5800      5810

    >< TthHB8I
    >< TaqI >< MaeIII
TTACACTACA ACCATCAAGC CTGTGTCGTA TAAACTCGAT GGAGTTACTT ACACAGAGAT TGAACCAAAA
    5820      5830      5840      5850      5860      5870      5880

    >< RsaI
    >< Csp6I
    >< SfcI >< BbvI
    >< Fnu4HI
    >< AfaI
TTGGATGGGT ATTATAAAAA GGATAATGCT TACTATACAG AGCAGCCTAT AGACCTTGTA CCAACTCAAC
    5890      5900      5910      5920      5930      5940      5950

    Tru9I ><
    SwaI ><
    MseI ><
    > < NspI
    > < NspHI
    > < NlaIII
    >< AflIII
CATTACCAAA TGCGAGTTTT GATAATTTCA AACTCACATG TTCTAACACA AAATTTGCTG ATGATTTAAA
    5960      5970      5980      5990      6000      6010      6020

    >< MboII
    >< AluI >< AluI >< MaeIII
TCAAATGACA GGCTTCACAA AGCCAGCTTC ACGAGAGCTA TCTGTCACAT TCTTCCCAGA CTTGAATGGC
    6030      6040      6050      6060      6070      6080      6090

    >< SfcI
GATGTAGTGG CTATTGACTA TAGACACTAT TCAGCGAGTT TCAAGAAAGG TGCTAAATTA CTGCATAAGC
    6100      6110      6120      6130      6140      6150      6160

    >< Tru9I
    >< ScrFI
    >< MvaI
    >< MseI
    >< EcoRII
    >< Ecl136I
    >< DsaV
    >< BstOI
    >< BstNI
    >< BsiLI
    >< MunI
    >< BstXI
    >< ApyI
    >< MaeII
    >< DraIII
    >< BstXI
CAATTGTTTG GCACATTAAC CAGGCTACAA CCAAGACAAC GTTCAAACCA AACACTTGGT GTTTACGTTG
    6170      6180      6190      6200      6210      6220      6230

    > < RsaI
    >< Csp6I
    > < AfaI >< BsrI
    >< MboII ><
    >< BbsI
TCTTTGGAGT ACAAAGCCAG TAGATACTTC AAATTCATTT GAAGTTCTGG CAGTAGAAGA CACACAAGGA
    6240      6250      6260      6270      6280      6290      6300

    >< HindII
    >< HincII
    >< MboII
    >< MnlI
    >< Eco57I
ATGGACAATC TTGCTTGTGA AAGTCAACAA CCCACCTCTG AAGAAGTAGT GGAAAATCCT ACCATACAGA
    6310      6320      6330      6340      6350      6360      6370

```

FIGURE 13.15

29/116

```

      >< MaeIII
      >< MaeII
AGGAAGTCAT AGAGTGTGAC GTGAAAAC TA CCGAAGTTGT AGGCAATGTC ATACTTAAAC CATCAGATGA
6380          6390          6400          6410          6420          6430          6440

      >< XhoII
      >< Sau3AI
      >< NlaIII
      >< NdeII
      >< MflI
      >< MboI
      >< DpnII
      >< DpnI
      >< BstYI
      >< BspAI
      >< Tru9I
      >< MseI
      >< BspHI >< Bsp143I>< Fnu4HI
      > < MaeIII >< MnlI >< BbvI >< AlwI
AGGTGTTAAA GTAACACAAG AGTTAGGTCA TGAGGATCTT ATGGCTGCTT ATGTGGAAAA CACAAGCATT
6450          6460          6470          6480          6490          6500          6510

      >< SauI
      >< RmaI
      >< MstII
      >< MaeI
      >< Eco81I
      >< DdeI
      >< CvnI
      >< Bsu36I
      >< Bse21I
      >< BfrI> < Tru9I
      >< AxyI> < MseI>< MunI >< NlaIII
      >< Tru9I >< MseI >< AluI >< AocI >< DraI >< BbvI Fnu4HI ><
ACCATTAAGA AACCTAATGA GCTTTCAC TA GCCTTAGGTT TAAAAACAAT TGCCACTCAT GGTATTGCTG
6520          6530          6540          6550          6560          6570          6580

      >< VspI >< StyI
      >< Tru9I >< EcoT14I > < DdeI
      >< MseI >< Eco130I >< BslI
      >< AsnI >< BssT1I >< BsiYI
      >< AseI >< BsaJI > < BfrI >< Fnu4HI
CAATTAATAG TGTTCCTTGG AGTAAAATTT TGGCTTATGT CAAACCATT C TTAGGACAAG CAGCAATTAC
6590          6600          6610          6620          6630          6640          6650

      >< HinPII
      >< Hin6I >< Tru9I
      >< HhaI >< MaeII>< MseI
      >< DdeI >< DraIII
      >< BbvI >< CfoI >< AflIII
AACATCAAAT TGCCTAAGA GATTAGCACA ACGTGTGTTT AACAATTATA TGCCTTATGT GTTTACATTA
6660          6670          6680          6690          6700          6710          6720

      >< RsaI > < RsaI>< XbaI
      >< Csp6I >< Csp6I >< RmaI
      >< MunI >< AfaI > < AfaI >< MaeI >< AluI
TTGTTCCAAT TGTGTACTTT TACTAAAAGT ACCAATTCTA GAATTAGAGC TTCACTACCT ACAACTATTG
6730          6740          6750          6760          6770          6780          6790

      >< VspI
      >< Tru9I
      >< NaeI
      >< MspI
      >< MseI

```

FIGURE 13. 16

30/116

```

                                >< HpaII
                                >< HapII
                                >< Cfr10I >< FokI
                                >< AsnI
                                >< AseI>< HphI>< MaeIII
                                >< Tru9I
                                >< MseI
                                >< SfaNI
CTAAAAATAG TGTTAAGAGT GTTGCTAAAT TATGTTTGGG TGCCGGCATT AATTATGTGA AGTCACCCAA
6800      6810      6820      6830      6840      6850      6860

                                >< Tru9I
                                >< MseI
                                >< DdeI
                                >< BfrI
                                >< MaeIII >
                                >< BbvI
ATTTTCTAAA TTGTTACAAA TCGCTATGTG GCTATTGTTG TTAAGTATTT GCTTAGGTTC TCTAATCTGT
6870      6880      6890      6900      6910      6920      6930

                                >< SduI
                                >< NspII
                                >< HgiAI
                                >< Bsp1286I
                                >< BmyI
                                >< Alw21I
                                >< RsaI
                                >< Csp6I
                                >< Fnu4HI
                                >< AfaI
GTAAGTGCTG CTTTGGTGT ACTCTTATCT AATTTTGGTG CTCCTTCTTA TTGTAATGGC GTTAGAGAAT
6940      6950      6960      6970      6980      6990      7000

                                >< Tru9I ><
                                >< MseI ><
                                >< Fnu4HI
                                >< MaeIII
                                >< MaeII
                                >< BbvI >
TGTATCTTAA TTCGTCTAAC GTTACTACTA TGGATTCTG TGAAGGTTCT TTTCCTTGCA GCATTTGTTT
7010      7020      7030      7040      7050      7060      7070

                                >< TfiI
                                >< MamI
                                >< HinfI
                                >< BsiBI
                                >< XmnI>< MaeIII
                                >< AluI >
                                >< PleI>< HinfI
                                >< BsaBI >< AluI
                                >< Asp700I
                                >< AfaI ><
AAGTGGATTA GACTCCCTTG ATTCTTATCC AGCTCTTGAA ACCATTGAGG TGACGATTTC ATCGTACAAG
7080      7090      7100      7110      7120      7130      7140

                                >< Pali
                                >< NspBII
                                >< HaeIII
                                >< GdiII
                                >< Fnu4HI
                                >< EaeI
                                >< DdeI
                                >< BsuRI
                                >< RmaI
                                >< BshI >< BslI
                                >< MaeI
                                >< AciI>< BsiYI
CTAGACTTGA CAATTTTAGG TCTGGCCGCT GAGTGGGTTT TGGCATATAT GTTGTTTACA AAATTCCTTT
7150      7160      7170      7180      7190      7200      7210

                                >< BspMI
                                >< RmaI
                                >< AluI
                                >< MaeI
ATTTATTAGG TCTTTCAGCT ATAATGCAGG TGTTCTTTGG CTATTTTGCT AGTCATTTC TCAGCAATTC
7220      7230      7240      7250      7260      7270      7280

                                >< RsaI ><
                                >< MboII
                                >< NlaIV
                                >< Eco64I
                                >< MamI ><
                                >< Csp6I ><
                                >< BscBI
                                >< BsiBI ><
                                >< BsaBI ><
                                >< AfaI ><
                                >< Csp6I >< BanI
                                >< AfaI>< AccBII
                                >< NlaIII

```

FIGURE 13.17

31/116

```

TTGGCTCATG TGGTTTATCA TTAGTATTGT ACAAATGGCA CCCGTTTCTG CAATGGTTAG GATGTACATC
  7290      7300      7310      7320      7330      7340      7350

                                TthHB8I ><
                                >< TaqI
                                MnlI ><
                                Ksp632I ><
                                >< FokI
                                >< EarI
                                >< MboII EarI ><
>< FokI                                >< Eam1104I>< AluI>< MboII >< NlaIII Eam1104I ><
TTCTTTGCTT CTTTCTACTA CATATGGAAG AGCTATGTTT ATATCATGGA TGGTTGCACC TCTTCGACTT
  7360      7370      7380      7390      7400      7410      7420

                                XhoII ><
                                Sau3AI ><
                                NlaIII ><
                                NdeII ><
                                MflI ><
                                MboI ><
                                >< ThaI
                                >< MvnI
                                >< EarI
                                >< Eam1104I
                                DpnII ><
                                BstYI ><
                                >< Bsp50I >< RsaI
                                >< AflIII >< Csp6I
                                >< AccII >< AfaI
                                >< MseI BglII ><
GCATGATGTG CTATAAGCGC AATCGTGCCA CACGCGTTGA GTGTACAAC ATTGTTAATG GCATGAAGAG
  7430      7440      7450      7460      7470      7480      7490

                                >< Pali
                                >< HaeIII
                                >< DsaI
                                >< MunI
                                >< MboII
                                >< BsuRI
                                >< BshI
                                >< MunI
                                >< BsmAI ><
>< DpnI                                >< Alw26I ><
>< Bsp143I >< MnlI >< BsaJI >< PleI>< HinfI
ATCTTTCTAT GTCTATGCAA ATGGAGGCCG TGGCTTCTGC AAGACTCACA ATTGGAATTG TCTCAATTGT
  7500      7510      7520      7530      7540      7550      7560

                                >< RsaI
                                >< Tru9I ><
                                >< Csp6I
                                >< MseI ><
                                >< BsrI
                                >< GsuI >< MaeIIIDraI ><
                                >< AfaI
                                >< BpmI
                                >< BsrI
GACACATTTT GCACTGGTAG TACATTCAAT AGTGATGAAG TTGCTCGTGA TTTGTCACTC CAGTTTAAAA
  7570      7580      7590      7600      7610      7620      7630

                                >< ThaI
                                >< MvnI
                                >< HphI
                                HinPlI ><
                                >< HinPlI
                                >< Hin6I
                                >< Hin6I
                                HhaI ><
                                >< HhaI
                                CfoI ><
                                >< CfoI
                                >< BstUI
                                >< BssHII
                                Bsp50I ><
                                >< BsrI
                                >< AccII
GACCAATCAA CCCTACTGAC CAGTCATCGT ATATTGTTGA TAGTGTGCT GTGAAAAATG GCGCGCTTCA
  7640      7650      7660      7670      7680      7690      7700

```

FIGURE 13. 18

32/116

```

                >< FokI
                    >< BsmAI
                >< MnlI                >< Alw26I                >< AciI
CCTCTACTTT GACAAGGCTG GTCAAAAGAC CTATGAGAGA CATCCGCTCT CCCATTTTGT CAATTTAGAC
    7710          7720          7730          7740          7750          7760          7770

                >< VspI
                >< Tru9I
                >< MseI
                >< AsnI
                >< AseI
                >< BcgI/a
    > < AluI
AATTTGAGAG CTAACAACAC TAAAGGTTCA CTGCCTATTA ATGTCATAGT TTTTGATGGC AAGTCCAAAT
    7780          7790          7800          7810          7820          7830          7840

                >< SfcI                >< PvuII
                >< RsaI                >< Psp5I
                >< Csp6I                >< NspBII
                >< PleI
    >< HinfI    >< DdeI    >< BcgI    >< AfaI    >< AluI
GCGACGAGTC TGCTTCTAAG TCTGCTTCTG TGTACTACAG TCAGCTGATG TGCCAACCTA TTCTGTTGCT
    7850          7860          7870          7880          7890          7900          7910

                TthHB8I ><
                TaqI ><
                SalI ><
                RtrI ><
                HindII >
                HincII >
                >< ScaI
                >< RsaI                >< Tru9I                HincII >
                >< Csp6I                >< SfaNI >< Eco57I
                >< AluI                >< AfaI                >< MseI                AccI ><
TGACCAAGCT CTTGTATCAG ACGTTGGAGA TAGTACTGAA GTTTCGGTTA AGATGTTTGA TGCTTATGTC
    7920          7930          7940          7950          7960          7970          7980

                >< Tru9I
                >< MseI
                > < Esp4I                >< SfcI
                > < AflIII                >< BspWI    >< AluI
GACACCTTTT CAGCAACTTT TAGTGTTCCCT ATGGAAAAAC TTAAGGCACT TGTTGCTACA GCTCACAGCG
    7990          8000          8010          8020          8030          8040          8050

                >< PvuII
                >< Psp5I
                >< NspBII
                >< Fnu4HI
                >< AluI                >< BbvI                >< AluI
AGTTAGCAAA GGGTGTAGCT TTAGATGGTG TCCTTTCTAC ATTCGTGTCA GCTGCCCCGAC AAGGTGTTGT
    8060          8070          8080          8090          8100          8110          8120

                >< HindII                >< BsmAI                MaeIII ><
                >< HincII                >< FokI>< Alw26I                >< DdeI
                >< BfrI
TGATACCGAT GTTGACACAA AGGATGTTAT TGAATGTCTC AAACCTTTCAC ATCACTCTGA CTTAGAAGTG
    8130          8140          8150          8160          8170          8180          8190

                >< XhoII
                Sau3AI ><
                >< NdeII
                >< MflI
                >< MboI
                >< NlaIII >< HgaI
                >< HinfI >< DpnII
                DpnI ><

```

FIGURE 13.19

```

                                Bsp143I ><
                                >< BsaHI >< BstYI
                                >< BbiIII >< BspAI
                                >< AcyI >< BglIII
                                >< MaeIII>< HphI
                                >< MaeIII >< HphI >< NlaIII
ACAGGTGACA GTTGTAAACA TTTCATGCTC ACCTATAATA AGGTTGAAAA CATGACGCCC AGAGATCTTG
      8200      8210      8220      8230      8240      8250      8260

      >< NspI
      >< NspHI
      >< NlaIII
>< HinPII
>< Hin6I
  >< HhaI
  >< CfoI
                                >< BspWI >< MaeIII
GCGCATGTAT TGA CTGTAA T GCAAGGCATA TCAATGCCCA AGTAGCAAAA AGTCACAATG TTTCACATCAT
      8270      8280      8290      8300      8310      8320      8330

                                >< NspI
                                >< NspHI
                                >< NlaIII
                                >< PvuII
                                >< Psp5I
                                >< NspBII
                                >< Fnu4HI
                                >< Eam1105I
                                >< BbvI
                                >< AflIII
                                >< AluI >< BbvI >< Fnu4HI
CTGGAATGTA AAAGACTACA TGTCTTTATC TGAACAGCTG CGTAAACAAA TTCGTAGTGC TGCCAAGAAG
      8340      8350      8360      8370      8380      8390      8400

                                >< RmaI
                                >< MaeI >< Eam1105I
AACAAACATAC CTTTTAGACT AACTTGTGCT ACAACTAGAC AGGTTGTCAA TGTCATAACT ACTAAAATCT
      8410      8420      8430      8440      8450      8460      8470

                                >< Tru9I
                                >< Pali
                                >< MseI
                                >< HaeIII
                                >< ScaI
                                >< Esp4I
                                >< RsaI >< Tru9I
                                >< Csp6I >< MseI
                                >< AfaI >< DraI >< AflIII >< BbvI
CACTCAAGGG TGGTAAGATT GTTAGTACTT GTTTTAAACT TATGCTTAAG GCCACATTAT TGTGCGTTCT
      8480      8490      8500      8510      8520      8530      8540

                                >< RsaI
                                >< Csp6I
                                >< BsrI
                                >< NlaIII
                                >< Fnu4HI
                                >< AfaI
                                >< MaeIII
TGCTGCATTG GTTTGTTATA TCGTTATGCC AGTACATACA TTGTCAATCC ATGATGGTTA CACAAATGAA
      8550      8560      8570      8580      8590      8600      8610

                                >< MaeIII
                                >< MaeIII
                                >< FokI
ATCATTTGGTT ACAAAGCCAT TCAGGATGGT GTCACCTCGTG ACATCATTTT TACTGATGAT TGTTTTGCAA
      8620      8630      8640      8650      8660      8670      8680

                                SfcI >
                                Fnu4HI ><
                                BbvI ><
                                >< NspI
                                >< NspHI
                                >< NlaIII
                                >< HgaI
                                >< BstXI
                                >< BbvI
                                >< AluI
ATAAACATGC TGGTTTTGAC GCATGGTTTA GCCAGCGTGG TGGTTCATAC AAAAATGACA AAAGCTGCCC
      8690      8700      8710      8720      8730      8740      8750

```

FIGURE 13. 20

34/116

```

                                >< ScrFI
                                >< ScrFI >< RsaI
                                >< MvaI >< MspI
                                >< EcoRII >< HpaII
                                >< Ecl136I>< NciI
                                >< DsaV >< HapII
                                >< BstOI>< DsaV
                                >< BstNI >< Csp6I
                                >< BsiLI >< BcnIDdeI ><
                                >< ApyI >< AfaI
                                >< Fnu4HI
                                >< AluI
TGTAGTAGCT GCTATCATTA CAAGAGAGAT TGGTTTCATA GTGCCTGGCT TACCGGGTAC TGTGCTGAGA
      8760      8770      8780      8790      8800      8810      8820

                                > < MaeIII >< HphI >< MnlI >< BspWI
GCAATCAATG GTGACTTCTT GCATTTTCTA CCTCGTGTTC TTAGTGCTGT TGGCAACATT TGCTACACAC
      8830      8840      8850      8860      8870      8880      8890

                                Tru9I >
                                SfaNI ><
                                >< RsaI
                                MseI >
                                >< BspWI >< Fnu4HI >< Csp6I
                                >< BbvI>< MnlI >< DdeI >< AfaI
CTTCCAACT CATTGAGTAT AGTGATTTTG CTACCTCTGC TTGCGTTCTT GCTGCTGAGT GTACAATTTT
      8900      8910      8920      8930      8940      8950      8960

                                > < RmaI
                                >< MnlI
                                > < MaeI
TAAGGATGCT ATGGGCAAAC CTGTGCCATA TTGTTATGAC ACTAATTTGC TAGAGGGTTC TATTTCTTAT
      8970      8980      8990      9000      9010      9020      9030

                                ScrFI >
                                MvaI >
                                MnlI ><
                                EcoRII ><
                                Ecl136I >
                                DsaV ><
                                BstOI >
                                >< NlaIV BstNI >
                                >< FokI BsiLI >
                                >< BscBI ApyI >
                                >< AluI
AGTGAGCTTC GTCCAGACAC TCGTTATGTG CTTATGGATG GTTCCATCAT ACAGTTTCCT AACACTTACC
      9040      9050      9060      9070      9080      9090      9100

                                >< RsaI
                                >< SfcI >< NspI
                                >< ScaI >< NspHI
                                >< RsaI >< NlaIII
                                >< MaeIII >< Csp6I >< NlaIII
                                >< GsuI >< AfaI >< Csp6I
                                >< BpmI >< DdeI >< AccI >< AfaI
TGGAGGGTTC TGTTAGAGTA GTAACAACTT TTGATGCTGA GTACTGTAGA CATGGTACAT GCGAAAGGTC
      9110      9120      9130      9140      9150      9160      9170

                                >< SstI
                                >< SduI
                                >< SacI
                                NspII ><
                                HgiAI ><
                                Eco24I ><
                                Bsp1286I ><

```

FIGURE 13.21

35/116

```

                                Ecl136II ><>< BmyI
                                BanII ><
                                >< Tru9I
                                >< MseI
                                Alw21I ><
                                >< BsrI
                                >< AluI
AGAAGTAGGT ATTTGCCTAT CTACCAGTGG TAGATGGGTT CTTAATAATG AGCATTACAG AGCTCTATCA
  9180          9190          9200          9210          9220          9230          9240

                                >< TfiI
                                >< SfaNI
                                >< HinfI
                                >< AluI
                                >< MnlI
GGAGTTTTCT GTGGTGTGGA TGCGATGAAT CTCATAGCTA ACATCTTTAC TCCTCTTG TG CAACCTGTGG
  9250          9260          9270          9280          9290          9300          9310

                                >< MaeIII
                                HphI ><
                                >< Eco57I
                                >< BbvI Fnu4HI ><
GTGCTTTAGA TGTGTCTGCT TCAGTAGTGG CTGGTGGTAT TATTGCCATA TTGGTGACTT GTGCTGCCTA
  9320          9330          9340          9350          9360          9370          9380

                                >< RsaI
                                >< Csp6I
                                >< NlaIII
                                >< MaeII
                                >< BbvI
                                >< Fnu4HI
                                >< AflIII
                                >< AfaI>< HphI
                                >< BspWI
CTACTTTATG AAATTCAGAC GTGTTTTTGG TGAGTACAAC CATGTTGTTG CTGCTAATGC ACTTTTGT TT
  9390          9400          9410          9420          9430          9440          9450

                                >< RsaI
                                >< NlaIV
                                >< KpnI
                                >< Eco64I
                                >< Csp6I
                                >< BscBI
                                >< Asp718
                                >< BanI >< AluI
                                >< AfaI
                                >< AccBI
                                >< Acc65I
                                >< HphI
                                >< NciI
                                >< MspI
                                >< HpaII
                                >< HinfI
                                >< HapII
                                >< PleI
                                >< BcnI
                                >< DdeI
                                >< AluI>< DsaV
                                >< AccI
TTGATGTCTT TCACTATACT CTGTCTGGTA CCAGCTTACA GCTTTCTGCC GGGAGTCTAC TCAGTCTTTT
  9460          9470          9480          9490          9500          9510          9520

                                >< RsaI
                                >< Csp6I
                                >< AfaI
                                >< HphI
                                >< HphI
                                NlaIII ><
ACTTGTA CTG GACATTCTAT TTCACCAATG ATGTTTCATT CTTGGCTCAC CTTCAATGGT TTGCCATGTT
  9530          9540          9550          9560          9570          9580          9590

TTCTCCTATT GTGCCTTTTT GGATAACAGC AATCTATGTA TTCTGTATTT CTCTGAAGCA CTGCCATTGG
  9600          9610          9620          9630          9640          9650          9660

                                >< TthHB8I
                                >< RsaI
                                >< MnlI
                                >< MnlI
                                >< Csp6I
                                >< Tru9I
                                >< PleI
                                >< BcgI/a
                                >< TaqI
                                >< MseI
                                >< DdeI
                                >< NlaIII
                                >< BbvI
                                >< Eco57I
                                >< BfrI
                                >< HinfI
                                >< MseI
                                >< MaeIII
                                >< AfaI
                                Fnu4HI ><
TTCTTTAACA ACTATCTTAG GAAAAGAGTC ATGTTTAATG GAGTTACATT TAGTACCTTC GAGGAGGCTG
  9670          9680          9690          9700          9710          9720          9730

                                >< RsaI
                                >< Csp6I
                                >< BcgI
                                >< RsaI
                                >< Csp6I
                                >< BsmAI

```

FIGURE 13.22

```

    >< AfaI
CTTTGTGTAC CTTTTTGCTC AACAAGGAAA TGTACCTAAA ATTGCGTAGC GAGACACTGT TGCCACTTAC
    9740          9750          9760          9770          9780          9790          9800

                                >< AfaI    >< Alw26I
                                >< RsaI      >< DdeI
                                >< Csp6I      >< BscBI
                                >< AfaI      >< BfrI    AluI ><
ACAGTATAAC AGGTATCTTG CTCTATATAA CAAGTACAAG TATTTTCAGTG GAGCCTTAGA TACTACCAGC
    9810          9820          9830          9840          9850          9860          9870

                                >< Fnu4HI
                                >< DdeI
                                >< Fnu4HI    >< BfrI
    >< BbvI    >< AluI    >< BbvI
TATCGTGAAG CAGCTTGCTG CCACTTAGCA AAGGCTCTAA ATGACTTTAG CAACTCAGGT GCTGATGTTC
    9880          9890          9900          9910          9920          9930          9940

                                >< SfcI
                                >< PstI      >< BsmI
                                >< BscCI
TCTACCAACC ACCACAGACA TCAATCACTT CTGCTGTTCT GCAGAGTGGT TTTAGGAAAA TGGCATTCCC
    9950          9960          9970          9980          9990          10000          10010

                                >< RsaI
                                >< NlaIII
                                >< MaeIII
                                >< Csp6I
                                >< AfaI
GTCAGGCAAA GTTGAAGGGT GCATGGTACA AGTAACCTGT GGAAGTACAA CTCTTAATGG ATTGTGGTTG
    10020          10030          10040          10050          10060          10070          10080

                                XhoII ><
                                Sau3AI ><
                                >< Tru9I    NdeII ><
                                >< NspI      MflI ><
                                >< NspHI     MboI ><
                                >< NlaIII    DpnII ><
                                >< MseI      BstYI ><
                                >< MboII    BspAI ><
                                >< AccI      AflIII >< BbsI    BglII ><
GATGACACAG TATACTGTCC AAGACATGTC ATTTGCACAG CAGAAGACAT GCTTAATCCT AACTATGAAG
    10090          10100          10110          10120          10130          10140          10150

                                Pali >
                                MscI >
                                HaeIII >
                                EaeI ><
                                BsuRI >
                                BshI >
                                BclI >
    >< DpnI >< MboII
    >< Bsp143I
ATCTGCTCAT TCGCAAATCC AACCATAGCT TTCTTGTTCA GGCTGGCAAT GTTCAACTTC GTGTTATTGG
    10160          10170          10180          10190          10200          10210          10220

                                >< DdeI> < Tru9I
                                >< BfrI> < MseI
                                >< DdeI
CCATTCTATG CAAAATTGTC TGCTTAGGCT TAAAGTTGAT ACTTCTAACC CTAAGACACC CAAGTATAAA
    10230          10240          10250          10260          10270          10280          10290

                                >< ScrFI
                                >< MvaI
    >< EcoRII
                                >< Ecl136I
                                >< SphI

```

FIGURE 13.23

37/116

```

    >< DsaV
    >< BstOI
    >< BstNI
    >< BsiLI
    >< ApyI
    TTTGTCCGTA TCCAACCTGG TCAAACATTT TCAGTTCTAG CATGCTACAA TGGTTCACCA TCTGGTGT TT
    10300      10310      10320      10330      10340      10350      10360

    >< PaeI
    >< NspI
    >< NspHI
    >< RmaI >< NlaIII
    >< MaeI >< HphI
    >< Sau3AI
    >< NdeII
    >< MboI>< NlaIII
    >< DpnII
    >< Eco31I
    >< BsmAI
    >< BsaI>< NlaIII
    >< Alw26I
    ATCAGTGTGC CATGAGACCT AATCATACCA TTAAAGGTTC TTTCCTTAAT GGATCATGTG GTAGTGTGTTG
    10370      10380      10390      10400      10410      10420      10430

    >< Tru9I
    >< MseI
    >< Zsp2I
    >< Ppu10I
    >< NsiI>< SfaNI
    >< NdeI
    >< Mph1103I
    >< EcoT22I
    >< AvaIII >< AluI >< RsaI ><
    >< Csp6I ><
    TTTTAACATT GATTATGATT GCGTGTCTTT CTGCTATATG CATCATATGG AGCTTCCAAC AGGAGTACAC
    10440      10450      10460      10470      10480      10490      10500

    >< SinI
    >< Sau96I
    >< NspIV
    >< NspHII
    >< Eco47I
    >< Cfr13I
    >< Bsi2I
    >< Bme18I >< HindII
    >< AvaII >< HincII
    >< AsuI>< BsgI >< BbvI >< BspMI AfaI ><
    GCTGGTACTG ACTTAGAAGG TAAATTCTAT GGTCCATTG TTGACAGACA AACTGCACAG GCTGCAGGTA
    10510      10520      10530      10540      10550      10560      10570

    >< Tru9I
    >< MseI
    >< BbvI
    >< Fnu4HI
    >< HphI ><
    CAGACACAAC CATAACATTA AATGTTTTTG CATGGCTGTA TGCTGCTGTT ATCAATGGTG ATAGGTGGTT
    10580      10590      10600      10610      10620      10630      10640

    >< Tru9I
    >< TfiI
    >< MseI
    >< HphI
    >< HinfI
    >< Tru9I
    >< MseI
    >< RsaI
    >< Csp6I
    >< AfaI
    TCTTAATAGA TTCACCACTA CTTTGAATGA CTTTAACCTT GTGGCAATGA AGTACAACCTA TGAACCTTTG
    10650      10660      10670      10680      10690      10700      10710

    >< SinI
    >< Sau96I
    >< PssI
    >< Psp5II
    >< PpuMI
    >< NspIV
    >< NspHII
    >< NlaIV

```

FIGURE 13. 24

38/116

```

                >< EcoO109I
                >< Eco47I
    >< Sau3AI      >< DraII
    >< NdeII       >< Cfr13I
    >< MboI        >< BsiZI
    >< DpnII>< NlaIII >< BscBI
    >< DpnI >< HindII >< Bme18I
    >< BspAI >< HincII >< AvaII
    >< Bsp143I    >< AsuI      >< MnlI
ACACAAGATC ATGTTGACAT ATTGGGACCT CTTTCTGCTC AAACAGGAAT TGCCGTCTTA GATATGTGTG
10720      10730      10740      10750      10760      10770      10780

                >< StyI
                >< RsaI
                >< EcoT14I
                >< Eco130I
    >< SfcI
    >< Fnu4HI      >< Fnu4HI
    >< BbvI        >< Fnu4HI
    >< BbvI        >< AluI >< PstI
CTGCTTTGAA AGAGCTGCTG CAGAATGGTA TGAATGGTCG TACTATCCTT GGTAGCACTA TTTTAGAAGA
10790      10800      10810      10820      10830      10840      10850

                >< StyI
                >< EcoT14I
                >< Eco130I
                >< BssT1I
    >< MboII
    >< MaeIII>< BsaJI
TGAGTTTACA CCATTTGATG TTGTTAGACA ATGCTCTGGT GTTACCTTCC AAGGTAAGTT CAAGAAAATT
10860      10870      10880      10890      10900      10910      10920

    >< SfaNI
    >< SduI
    >< NspII
    >< Tru9I>< Bsp1286I >< Tru9I
    >< MseI >< BmyI      >< MseI
GTTAAGGGCA CTCATCATTG GATGCTTTTA ACTTTCTTGA CATCACTATT GATTCTTGTT CAAAGTACAC
10930      10940      10950      10960      10970      10980      10990

                >< XmnI
                >< BsmI
                >< BscCI
    >< MaeIII
    >< Asp700I
AGTGGTCACT GTTTTTCTTT GTTTACGAGA ATGCTTTCTT GCCATTTACT CTTGGTATTA TGGCAATTGC
11000      11010      11020      11030      11040      11050      11060

    >< NspI
    >< NspHI
    >< NlaIII
    >< BspWI >< Fnu4HI>< BspWI >< BscCI
    >< MaeIII
TGCATGTGCT ATGCTGCTTG TTAAGCATAA GCACGCATTC TTGTGCTTGT TTCTGTTACC TTCTCTTGCA
11070      11080      11090      11100      11110      11120      11130

                >< SfaNI
                >< RmaI
    >< NspI
    >< NlaIII
    >< NheI
    >< Tru9I
    >< BspWI >< MseI >< AccI>< NspHI>< AluI
    >< BsiBI >< NlaIII
ACAGTTGCTT ACTTTAATAT GGTCTACATG CCTGCTAGCT GGGTGATGCG TATCATGACA TGGCTTGAAT
11140      11150      11160      11170      11180      11190      11200

```

FIGURE 13.25

39/116

```

                                >< Tru9I
                                >< MseI
    > < RmaI                    > < Esp4I
    > < MaeI                    >< Eco57I
                                >< AluI
    >< AluI                    > < AflII
TGGCTGACAC TAGCTTGTCT GGTATAGGC TTAAGGATTG TGTTATGTAT GCTTCAGCTT TAGTTTTGCT
    11210      11220      11230      11240      11250      11260      11270

                                >< RmaI
                                >< MaeII
                                >< MaeI
    > < NlaIII    >< SfaNI    >< Fnu4HI
    >< BspHI >< AluI    >< BbvI    >< AflIII
TATTCTCATG ACAGCTCGCA CTGTTTATGA TGATGCTGCT AGACGTGTTT GGACACTGAT GAATGTCATT
    11280      11290      11300      11310      11320      11330      11340

                                >< Sau96I
                                >< Pali
                                >< NspIV
    >< NlaIII
                                >< HaeIII
                                > < DdeI
                                >< Cfr13I
                                >< BsuRI
                                >< BsiZI
                                >< BshI
                                > < BfrI
                                >< AsuI
    >< AccI
    >< BspAI>< AluI
ACACTTGTTT ACAAAGTCTA CTATGGTAAT GCTTTAGATC AAGCTATTTT CATGTGGGCC TTAGTTATTT
    11350      11360      11370      11380      11390      11400      11410

                                >< RmaI
                                >< NlaIII
                                >< MaeI>< SfcI
                                >< AluI>< AluI
    >< MaeIII    >< MnlI    >< MaeIII
    >< MaeIII    >< MnlI    >< MaeIII
CTGTAACCTC TAACTATTCT GGTGTCGTTA CGACTATCAT GTTTTTAGCT AGAGCTATAG TGTTTGTGTG
    11420      11430      11440      11450      11460      11470      11480

                                DdeI >
                                >< NlaIII BfrI >
    >< BsrI
    >< NlaIII BfrI >
TGTTGAGTAT TACCCATTGT TATTTATTAC TGGCAACACC TTACAGTGTA TCATGCTTGT TTATTGTTTC
    11490      11500      11510      11520      11530      11540      11550

                                >< Pali
                                >< HaeIII
                                >< Fnu4HI    >< BsuRI
    >< BbvI    >< Fnu4HI    >< BspWI
    >< BbvI    >< BspWI    >< BshI    >< Eco57I    >< MaeIII
TTAGGCTATT GTTGCTGCTG CTACTTTGGC CTTTTCTGTT TACTCAACCG TTA CTACTCTTG
    11560      11570      11580      11590      11600      11610      11620

                                >< ScrFI
                                >< MvaI
                                >< EcoRII
                                >< Ecl136I
                                >< DsaV
                                >< BstOI
                                >< BstNI
                                >< BsiLI
                                > < BsaJI
                                >< BsaJI
    >< Eco31I
    >< BsmAI
    >< BsaI

```

FIGURE 13. 26

40/116

```

      >< DrdI  >< Alw26I
GTGTTTATGA CTACTTGGTC TCTACACAAG AATTTAGGTA TATGAACTCC CAGGGGCTTT TGCCTCCTAA
11630      11640      11650      11660      11670      11680      11690

      >< Tru9I
      >< MseI
>< SfaNI      > < HindIII> < Tru9I
  >< MnlI      >< AluI > < MseI  > < MnlI      > < NlaIII
GAGTAGTATT GATGCTTTCA AGCTTAACAT TAAGTTGTTG GGTATTGGAG GTAAACCATG TATCAAGGTT
11700      11710      11720      11730      11740      11750      11760

      >< VneI
      >< SnoI
      >< SduI
      >< NspII
      >< HgiAI
      >< Bsp1286I
      >< BmyI      >< RsaI
      >< RsaI
      >< Csp6I      >< ApaLI      >< MboII
      >< AfaI      >< MaeII      >< Alw44I      >< Csp6I      DdeI >
GCTACTGTAC AGTCTAAAAT GTCTGACGTA AAGTGCACAT CTGTGGTACT GCTCTCGGTT CTTCAACAAC
11770      11780      11790      11800      11810      11820      11830

      >< NspII> < RsaI
      >< DraIII
      >< SduI>< Csp6I
      >< Bsp1286I
      >< MboII
      >< HinfI >< PleI      >< BmyI > < AfaI      >< MboII
TTAGAGTAGA GTCATCTTCT AAATTGTGGG CACAATGTGT ACAACTCCAC AATGATATTC TTCTTGCAAA
11840      11850      11860      11870      11880      11890      11900

      >< TthHB8I
      >< TaqI
      >< HindIII      >< MboII      SfcI ><
      >< AluI      > < Eco57I      >< NlaIII
AGACACAAC T GAAGCTTTTCG AGAAGATGGT TTCTCTTTTG TCTGTTTTGC TATCCATGCA GGGTGTGTA
11910      11920      11930      11940      11950      11960      11970

      >< VspI
      >< Tru9I
      >< MseI      >< TthHB8I      > < Ksp632I
      >< AsnI      >< TaqI >< MboII      > < EarI
      >< AseI>< MnlI >< BcgI/a      >< Eco57I      >< Eco57I >< BcgI
GACATTAATA GGTGTGCGA GGAAATGCTC GATAACCGTG CTACTCTTCA GGCTATTGCT TCAGAATTTA
11980      11990      12000      12010      12020      12030      12040

      >< StuI
      >< ScrFI
      >< Pali
      >< MvaI>< HaeIII
      >< EcoRII>< Eco147I
      >< Ecl136I
      >< DsaV >< BsuRI
      >< BstOI
      >< BstNI
      >< BspWI
      >< BsiLI
      >< Fnu4HI      >< BsaJI >< BshI
      >< NdeI      >< BspWI>< MnlI >< BglI      TfiI ><
      >< AciI      >< ApyI>< AatI      >< SfcI HinfI ><
      > < AluI

```

FIGURE 13. 27

41/116

```

GTTCTTTACC ATCATATGCC GCTTATGCCA CTGCCCAGGA GGCCTATGAG CAGGCTGTAG CTAATGGTGA
12050      12060      12070      12080      12090      12100      12110

    >< XmnI      >< Tru9I      >< SfaNI
    >< HphI      >< MseI      >< DdeI
    >< Asp700I   >< Eco57I   >< BbvI Fnu4HI ><
TTCTGAAGTC GTTCTCAAAA AGTTAAAGAA ATCTTTGAAT GTGGCTAAAT CTGAGTTTGA CCGTGATGCT
12120      12130      12140      12150      12160      12170      12180

                                XhoII ><
                                Sau3AI ><
                                NdeII ><
                                MnlI >
                                >< MnlI
                                >< MflI
                                >< MboI
                                DpnII ><
                                DpnI ><
                                DdeI ><
                                BstYI ><
                                >< BspWI      >< RsaIBspAI ><
                                >< BspAI      >< Csp6IBsp143I ><
                                >< Bsp143I   >< AfaIBglII ><
    >< NlaIII
GCCATGCAAC GCAAGTTGGA AAAGATGGCA GATCAGGCTA TGACCCAAAT GTACAAACAG GCAAGATCTG
12190      12200      12210      12220      12230      12240      12250

                                >< SpeI
                                >< RmaI
                                >< MaeIII      >< MboII      >< Eam1104I >< BspWI
                                >< MaeI      >< BspWI      >< EarI>< BfrI >< AluI
AGGACAAGAG GGCAAAAGTA ACTAGTGCTA TGCAAACAAT GCTCTTCACT ATGCTTAGGA AGCTTGATAA
12260      12270      12280      12290      12300      12310      12320

                                >< ThaI
                                >< MvnI
                                >< HinPII
                                >< Hin6I
                                >< HhaI
                                >< CfoI
                                >< BstUI
                                >< Bsp50I
                                >< AccII
                                SfcI ><
TGATGCACTT AACAAACATTA TCAACAATGC GCGTGATGGT TGTGTTCCAC TCAACATCAT ACCATTGACT
12330      12340      12350      12360      12370      12380      12390

                                >< RsaI
                                >< NlaIV
                                >< Eco64I
                                >< Csp6I
                                >< BslI
                                >< BsiYI>< KpnI
                                >< BscBI
                                >< BanI
                                >< Asp718
                                >< AfaI
                                >< AccB1I      >< MaeIII
                                >< Acc65I      BsgI ><
ACAGCAGCCA AACTCATGGT TGTTGTCCCT GATTATGGTA CCTACAAGAA CACTTGTGAT GGTAACACCT
12400      12410      12420      12430      12440      12450      12460

                                >< Zsp2I
                                >< Ppu10I

```

FIGURE 13.28


```

    >< NsiI
    >< Mph1103I
    >< NdeI>< EcoT22I
    >< AvaIII >< SfaNI
    >< SfaNI
    >< AciI
    DdeI ><
    BfrI ><
    TTACATATGC ATCTGCACTC TGGGAAATCC AGCAAGTTGT TGATGCGGAT AGCAAGATTG TTCAACTTAG
    12470 12480 12490 12500 12510 12520 12530

    >< Pali
    >< HaeIII >< MnlI >< DdeIDdeI ><
    >< BsuRI >< MaeIII >< BspWI
    >< Tru9I>< NlaIII
    >< MseI>< HphI > < XcmI>< BshI >< AluI BspWI ><
    TGAAATTAAC ATGGACAATT CACCAAATTT GGCTTGGCCT CTTATTGTTA CAGCTCTAAG AGCCAACTCA
    12540 12550 12560 12570 12580 12590 12600

    RsaI ><
    NlaIV ><
    KpnI ><
    >< Fnu4HI
    Eco64I ><
    Csp6I ><
    BscBI ><
    Asp718 ><
    AfaI ><
    >< AciI>< BanI
    AccBI ><
    >< MseI >< HinfI >< PliI
    >< DdeI>< BsrI >< PshAI Acc65I ><
    >< AluI > < SfcI >< DdeI>< BsrI >< PshAI Acc65I ><
    GCTGTTAAAC TACAGAATAA TGAAGTGAAGT CCAGTAGCAC TACGACAGAT GTCCTGTGCG GCTGGTACCA
    12610 12620 12630 12640 12650 12660 12670

    >< TthHB8I
    >< TaqI
    >< SfuI
    >< NspV
    >< MnlI
    >< LspI
    >< Csp45I
    >< BstBI
    >< Bsp119I
    >< BsiCI
    >< Bpu14I
    >< AsuII
    >< RsaI
    >< Csp6I
    >< AluI
    >< AfaI
    CACAAACAGC TTGTACTGAT GACAATGCAC TTGCCTACTA TAACAATTCTG AAGGGAGGTA GGTTCGTGCT
    12680 12690 12700 12710 12720 12730 12740

    >< XhoII
    >< Sau3AI
    >< NdeII
    >< MflI
    >< MboI
    >< DpnII
    >< DpnI
    >< BstYI >< TfiI >< RsaI
    >< BspAI >< RmaI >< Csp6I
    >< Bsp143I >< HinfI >< Csp6I>< RsaI
    >< BglII >< MaeI >< DdeI >< AfaI>< AfaI
    GGCATTACTA TCAGACCACC AAGATCTCAA ATGGGCTAGA TTCCCTAAGA GTGATGGTAC AGGTACAATT
    12750 12760 12770 12780 12790 12800 12810

    >< Sau96I
    >< PssI
    >< Pali
    >< NspIV

```

FIGURE 13.29

43/116

```

                                >< HaeIII
                                >< EcoO109I
                                >< DraII
                                >< Cfr13I
                                >< BsuRI
                                >< BsiZI      RsaI >
                                >< BshI      Csp6I ><
                                >< AsuI      AfaI >
TACACAGAAC TGGAAACCACC TTGTAGGTTT GTTACAGACA CACCAAAAGG GCCTAAAGTG AAATACTTGT
  12820      12830      12840      12850      12860      12870      12880

                                >< SfcI
                                > < MboII
                                MaeII ><
                                >< Fnu4HI >< RsaI
                                >< Eco57I >< Csp6I
                                > < BbsI
                                >< Tru9I
                                >< MseI >< MnlI
                                >< BbvI      >< AluI      >< AfaI
ACTTCATCAA AGGCTTAAAC AACCTAAATA GAGGTATGGT GCTGGGCAGT TTAGCTGCTA CAGTACGTCT
  12890      12900      12910      12920      12930      12940      12950

                                >< RsaI
                                >< SfcI >< Csp6I
                                >< BspWI      >< AfaI      >< BspMI
TCAGGCTGGA AATGCTACAG AAGTACCTGC CAATTCAACT GTGCTTTCCT TCTGTGCTTT TGCAGTAGAC
  12960      12970      12980      12990      13000      13010      13020

                                >< RmaI
                                >< MnlI
                                >< MaeI      >< HphI
CCTGCTAAAG CATATAAGGA TTACCTAGCA AGTGGAGGAC AACCAATCAC CAACTGTGTG AAGATGTTGT
  13030      13040      13050      13060      13070      13080      13090

                                >< SinI
                                >< Sau96I
                                >< NspIV
                                >< NspHII
                                >< NlaIII
                                >< Eco47I
                                >< Eam1105I
                                >< Cfr13I
                                >< BsiZI
                                >< Bme18I >< XcmI
                                >< AvaII   PleI ><
                                >< AsuI> < HinfI
GTACACACAC TGGTACAGGA CAGGCAATTA CTGTAACACC AGAAGCTAAC ATGGACCAAG AGTCCTTTGG
  13100      13110      13120      13130      13140      13150      13160

                                >< TfiI
                                >< MaeIII
                                >< HinfI
TGGTGCTTCA TGTTGTCTGT ATTGTAGATG CCACATTGAC CATCCAAATC CTAAAGGATT CTGTGACTTG
  13170      13180      13190      13200      13210      13220      13230

                                > < RsaI
                                >< MaeII
                                >< Csp6I
                                > < AfaI
                                >< BspMI
                                >< DdeI
                                >< BfrI
AAAGGTAAGT ACGTCCAAAT ACCTACCACT TGTGCTAATG ACCCAGTGGG TTTTACACTT AGAAACACAG
  13240      13250      13260      13270      13280      13290      13300

                                >< ThaI

```

FIGURE 13.30

44/116

```

                                >< SfaNI
                                >< MvnI
                                >< BstUI
                                >< Bsp50I
                                >< AciI
                                >< AccIISfaNI ><
>< RsaI
>< Csp6I
>< AfaI >< AciI
                                >< SfcI >< MaeIII
                                >< AccIISfaNI ><
TCTGTACCGT CTGCGGAATG TGGAAAGGTT ATGGCTGTAG TTGTGACCAA CTCCGCGAAC CCTTGATGCA
13310      13320      13330      13340      13350      13360      13370

                                >< Zsp2I
                                > < SfaNI
                                >< Mph1103I>< Tru9I
>< Ppu10I>< MaeII
                                Fnu4HI ><
                                BsgI ><
                                >< BbvI
                                >< AciI>< AvaIII >< DraI >< AciI >< Fnu4HI >< AciI ><
GTCTGCGGAT GCATCAACGT TTTTAAACGG GTTTCGCGTG TAAGTGCAGC CCGTCTTACA CCGTGC GGCA
13380      13390      13400      13410      13420      13430      13440

>< SpeI
>< ScaI
>< RsaI
>< RmaI
>< MaeI
> < Csp6I >< SfcI
                                >< BspWI
>< BspWI >< AfaI >< AccI >< BcgI/a
                                BcgI >
CAGGCACTAG TACTGATGTC GTCTACAGGG CTTTGTATAT TTACAACGAA AAAGTTGCTG GTTTTGCAAA
13450      13460      13470      13480      13490      13500      13510

                                >< ScrFI
                                >< MvaI
                                >< MnlI
                                >< EcoRII
                                >< Ecl136I
                                >< BstOI
                                >< BstNI
                                >< BslI
                                >< DsaV >< BsiYI
                                >< BsiLI
                                >< ApyI
                                >< PleI
                                > < FokI >< HinfI
GTTCTAAAAA ACTAATTGCT GTCGCTTCCA GGAGAAGGAT GAGGAAGGCA ATTTATTAGA CTCTTACTTT
13520      13530      13540      13550      13560      13570      13580

                                >< NlaIII
                                >< Ksp632I
                                >< EarI
                                >< Eam1104I
                                >< BsmAI
                                >< Tru9I
>< Tru9I
>< MseI
                                >< Alw26I
                                >< MboII >< MseI
>< MnlI
GTAGTTAAGA GGCATACTAT GTCTAACTAC CAACATGAAG AGACTATTTA TAACTTGGTT AAAGATTGTC
13590      13600      13610      13620      13630      13640      13650

                                >< RsaI
                                >< NlaIV
                                > < NlaIII
                                >< KpnI
                                >< HphI
                                > < Eco64I
                                >< Csp6I
                                >< BscBI
                                > < BanI
                                > < Asp718

```

FIGURE 13.31

```

                                >< MaeIII >< AfaI
                                > < AccBII MaeII ><
>< NspBII                                > < Acc65I > < HgaI
>< AciI                                >< NlaIII                                >< NlaIII
CAGCGGTTGC TGTCCATGAC TTTTCAAGT TTAGAGTAGA TGGTGACATG GTACCACATA TATCACGTCA
13660      13670      13680      13690      13700      13710      13720

                                >< MnlI
                                >< MaeII
GCGTCTAACT AAATACACAA TGGCTGATTT AGTCTATGCT CTACGTCATT TTGATGAGGG TAATTGTGAT
13730      13740      13750      13760      13770      13780      13790

>< Tru9I
>< MseI                                >< MaeIII >< MunI
ACATTAAAAG AAATACTCGT CACATACAAT TGCTGTGATG ATGATTATTT CAATAAGAAG GATTGGTATG
13800      13810      13820      13830      13840      13850      13860

                                >< ThaI
                                >< MvnI
                                >< MluI
                                >< BstUI
                                >< Bsp50I
                                >< RsaI
                                >< HphI
                                >< Csp6I Tru9I ><
                                >< AfaI MseI ><
>< TfiI                                >< AflIII                                >< DdeI
>< HinfI                                >< AccII                                >< BfrI
ACTTCGTAGA GAATCCTGAC ATCTTACGCG TATATGCTAA CTTAGGTGAG CGTGTACGCC AATCATTATT
13870      13880      13890      13900      13910      13920      13930

                                >< XhoII >
                                >< Sau3AI >
                                >< NdeII >
                                >< MflI >
                                >< MboI >
                                >< DpnII >
                                >< BstYI >
                                >< BspAI >
                                >< RsaI
                                >< Csp6I
                                >< BspWI
                                >< AfaI
> < SfaNI                                >< SfaNI
>< RsaI                                >< SfaNI
>< Csp6I                                >< SfaNI
>< AfaI                                >< SfaNI
AAAGACTGTA CAATTCTGCG ATGCTATGCG TGATGCAGGC ATTGTAGGCG TACTGACATT AGATAATCAG
13940      13950      13960      13970      13980      13990      14000

                                > < ScrFI
                                > < MvaI
                                >< Fnu4HI
                                >< EcoRII
                                > < Ecl136I
                                > < BstOI
                                > < BstNI
                                >< RsaI
                                >< BslI
                                > < HphI
                                >< Csp6I
                                > < BsiLI
                                > < BbvI
                                > < ApyI
                                >< AfaI
                                >< DsaV
                                >< AciI
>< Tru9I                                >< RsaI
>< MseI                                >< RsaI
>< DpnI                                >< Csp6I
>< Bsp143I                                >< BsrI
>< AlwI                                >< AfaI
GATCTTAATG GGAAGTGGTA CGATTTCGGT GATTTCGTAC AAGTAGCACC AGGCTGCGGA GTTCCTATTG
14010      14020      14030      14040      14050      14060      14070

                                >< SfaNI
                                >< RmaI
                                > < HinfI
                                >< MnlI
                                >< Fnu4HIpleI ><
                                >< MaeI
                                >< DdeI
>< TfiI                                >< SfaNI
>< HinfI                                >< FokI
>< MamI                                >< BsiBI
>< BsaBI                                >< BsaBI
TGGATTCATA TTAATCATTG CTGATGCCCA TCCTCACTTT GACTAGGGCA TTGGCTGCTG AGTCCCATAT
14080      14090      14100      14110      14120      14130      14140

>< Sau3AI
>< NdeII

```

FIGURE 13.32

46/116

```

>< MboI
>< MamI
>< DpnII
>< DpnI
>< BspWI
>< BspAI
>< Bsp143I
>< BsiBI
>< BsaBI >< FokI
GGATGCTGAT CTCGCAAAAC CACTTATTAA GTGGGATTG CTGAAATATG ATTTTACGGA AGAGAGACTT
14150      14160      14170      14180      14190      14200      14210

>< SinI
>< Sau96I
>< NspIV
>< NspHII
>< NlaIV
>< FokI
>< Eco47I
>< Cfr13I
>< Bsi2I
>< BscBI
>< Tru9I
>< Bme18I
>< AvaII
>< AsuI
>< MunI
>< MseI
TGTCTCTTCG ACCGTTATTT TAAATATTGG GACCAGACAT ACCATCCCAA TTGTATTAAC TGTTTGGATG
14220      14230      14240      14250      14260      14270      14280

>< SinI ><
>< Sau96I ><
>< NspIV ><
>< NspHII >
>< Eco47I ><
>< Cfr13I ><
>< Bsi2I ><
>< Bme18I ><
>< AvaII ><
>< AsuI ><
>< FokI
>< MseI
ATAGGTGTAT CCTTCATTGT GCAAACCTTA ATGTGTTATT TTCTACTGTG TTTCCACCTA CAAGTTTGGG
14290      14300      14310      14320      14330      14340      14350

>< SpeI
>< RmaI
>< MaeI
>< SspI
>< BsrI
ACCACTAGTA AGAAAAATAT TTGTAGATGG TGTTCCTTTT GTTGTTTCAA CTGGATACCA TTTTCGTGAG
14360      14370      14380      14390      14400      14410      14420

>< ThaI>< Esp3I
>< DdeI
>< BstUI
>< Bsp50I >< BsmBI
>< MvnI>< BsmAI
>< HgaI>< AluI >< Alw26I
>< FokI >< AccII >< BbvI
TTAGGAGTCG TACATAATCA GGATGTAAAC TTACATAGCT CGCGTCTCAG TTTCAAGGAA CTTTTAGTGT
14430      14440      14450      14460      14470      14480      14490

>< Zsp2I
>< SphI
>< Ppu10I
>< PaeI
>< NspI

```

FIGURE 13.33

```

    >< Sau3AI          >< NspHI
    >< NdeII           >< NsiI
    >< MboI             >< NlaIII
    >< DpnII           >< Mph1103I
    > < DpnI          >< Fnu4HI
    >< Fnu4HI>< BspWI  >< EcoT22I
    >< BspAI           >< BspWI
    > < Bsp143I> < AvaIII > < AlwNI
    >< AlwI           >< AluI   >< AluI   >< BbvI   >< MaeI
    ATGCTGCTGA TCCAGCTATG CATGCAGCTT CTGGCAATTT ATTGCTAGAT AAACGCACTA CATGCTTTTC
    14500      14510      14520      14530      14540      14550      14560

    >< ScrFI
    >< NciI
    >< MspI
    >< HpaII
    >< HapII
    >< Fnu4HI
    >< AlwNI
    >< AluI
    AGTAGCTGCA CTAACAAACA ATGTTGCTTT TCAAAGTGTG AAACCCGGTA ATTTTAATAA AGACTTTTAT
    14570      14580      14590      14600      14610      14620      14630

    >< Tru9I
    >< MseI
    GACTTTGCTG TGTCTAAAGG TTTCTTTAAG GAAGGAAGTT CTGTTGAAGT AAAACACTTC TTCTTTGCTC
    14640      14650      14660      14670      14680      14690      14700

    >< FokI
    >< Fnu4HI
    AGGATGGCAA CGCTGCTATC AGTGATTATG ACTATTATCG TTATAATCTG CCAACAATGT GTGATATCAG
    14710      14720      14730      14740      14750      14760      14770

    >< VspI
    >< Tru9I
    >< MseI
    >< AsnI
    >< AseI
    ACAACTCCTA TTCGTAGTTG AAGTTGTTGA TAAATACTTT GATTGTTACG ATGGTGGCTG TATTAATGCC
    14780      14790      14800      14810      14820      14830      14840

    >< Tru9I
    >< MseI
    >< HpaI
    >< HindII
    >< HincII
    AACCAAGTAA TCGTTAACAA TCTGGATAAA TCAGCTGGTT TCCCATTAA TAAATGGGGT AAGGCTAGAC
    14850      14860      14870      14880      14890      14900      14910

    >< SfaNI
    >< Sau3AI
    >< NdeII
    >< MboI
    >< DpnII
    >< DpnI
    >< Bsp143I
    >< PleI
    >< HinfI>< MnlI
    >< BspAI
    >< AlwI
    TTTATTATGA CTCAATGAGT TATGAGGATC AAGATGCACT TTTCGCGTAT ACTAAGCGTA ATGTCATCCC
    14920      14930      14940      14950      14960      14970      14980

    >< ThaI
    >< MvnI
    >< BstUI
    >< Bst1107I
    >< BspWI
    >< FokI
    >< Bsp50I
    >< AccII>< DdeI
    >< AccI

    >< SstI
    >< SduI
    >< SacI

```

FIGURE 13.34

48/116

```

                                >< NspII
                                >< HgiAI
                                >< Eco24I
                                > < Ecl136II
                                >< Bsp1286I
                                >< BmyI
                                >< BanII
                                >< Alw21I
                                > < AluI
                                >< AluI
TACTATAACT CAAATGAATC TTAAGTATGC CATTAGTGCA AAGAATAGAG CTCGCACCGT AGCTGGTGTG
14990      15000      15010      15020      15030      15040      15050

                                RmaI ><
                                > < MnlI
                                MaeI ><
                                >< Fnu4HI
                                >< AciI
TCTATCTGTA GTACTATGAC AAATAGACAG TTTCATCAGA AATTATTGAA GTCAATAGCC GCCACTAGAG
15060      15070      15080      15090      15100      15110      15120

                                >< Tru9I
                                >< MseI
GAGCTACTGT GGTAATTGGA ACAAGCAAGT TTTACGGTGG CTGGCATAAT ATGTTAAAAA CTGTTTACAG
15130      15140      15150      15160      15170      15180      15190

                                NspI ><
                                NspHI ><
                                NlaIII ><
                                >< NlaIII
                                DdeI ><
                                BspWI ><
                                >< MaeIII
                                BfrI ><
TGATGTAGAA ACTCCACACC TTATGGGTTG GGATTATCCA AAATGTGACA GAGCCATGCC TAACATGCTT
15200      15210      15220      15230      15240      15250      15260

                                > < Pali
                                > < HaeIII
                                > < BsuRI
                                > < BshI
                                >< MnlI
                                >< MaeIII
                                SfcI ><
AGGATAATGG CCTCTCTTGT TCTTGCTCGC AAACATAACA CTTGCTGTAA CTTATCACAC CGTTTCTACA
15270      15280      15290      15300      15310      15320      15330

                                Tru9I ><
                                ScrFI >
                                MvaI >
                                >< MseI
                                FokI ><
                                EcoRII ><
                                Ecl136I >
                                DsaV ><
                                BstOI >
                                BstNI >
                                >< NlaIII
                                > < Fnu4HI
                                BsiLI >
                                >< AciI
                                ApyI >
GGTTAGCTAA CGAGTGTGCG CAAGTATTAA GTGAGATGGT CATGTGTGGC GGCTCACTAT ATGTTAAACC
15340      15350      15360      15370      15380      15390      15400

                                > < SfaNI
                                >< MspI
                                >< HpaII
                                >< HapII
                                >< HphI
                                >< BspWI
                                >< Tru9I
                                MaeIII ><
                                >< MseI
                                AluI ><

```

FIGURE 13.35

49/116

```

AGGTGGAACA TCATCCGGTG ATGCTACAAC TGCTTATGCT AATAGTGTCT TTAACATTTG TCAAGCTGTT
15410      15420      15430      15440      15450      15460      15470

>< BspWI                                     >< DrdI
ACAGCCAATG TAAATGCACT TCTTTCAACT GATGGTAATA AGATAGCTGA CAAGTATGTC CGCAATCTAC
15480      15490      15500      15510      15520      15530      15540

                                     >< AluI       > < AciI
                                     >< Sau3AI
                                     >< NdeII
                                     >< MboI
> < MamI
  >< FbaI
>< DpnII
  >< DpnI
    >< BspHI
>< BspAI
  >< Bsp143I
>< BsiQI
      >< SfcI
      >< BsmAI
      >< Alw26I
      > < BsiBI>< NlaIII
      > < BsaBI>< FokI
      >< BclI>< EcoRI
AACACAGGCT CTATGAGTGT CTCTATAGAA ATAGGGATGT TGATCATGAA TTCGTGGATG AGTTTTACGC
15550      15560      15570      15580      15590      15600      15610
                                     FokI ><

      >< TfiI
      >< SfaNI
      >< NlaIII
      >< BspMI
      >< HinfI
      >< MaeIII
TTACCTGCGT AAACATTTCT CCATGATGAT TCTTTCTGAT GATGCCGTTG TGTGCTATAA CAGTAACTAT
15620      15630      15640      15650      15660      15670      15680

      > < RmaI
      >< NheI >< Tru9I
>< Fnu4HI      > < MaeI      >< Tru9I
>< AciI      >< AluI >< MseI      >< MseI
GCGGCTCAAG GTTTAGTAGC TAGCATTAAG AACTTTAAGG CAGTTCTTTA TTATCAAAAT AATGTGTTCA
15690      15700      15710      15720      15730      15740      15750
                                     MnlI ><

                                     >< SinI
                                     >< Sau96I
                                     >< PssI
                                     >< Psp5II
                                     >< PpuMI
                                     >< NspIV
                                     >< NspHII
                                     >< EcoO109I
                                     >< Eco47I
                                     >< DraII
                                     >< Cfr13I
                                     >< BsiZI
                                     >< Bme18I
                                     >< AvaII
                                     >< AsuI
                                     >< MnlI
      >< DdeI
      >< NlaIII
      >< DdeI
      >< BsmAI
      >< Alw26I
TGCTCTGAGGC AAAATGTTGG ACTGAGACTG ACCTTACTAA AGGACCTCAC GAATTTTGCT CACAGCATAAC
15760      15770      15780      15790      15800      15810      15820

                                     >< XhoII
                                     >< Sau3AI
                                     >< NdeII
                                     >< MflI
                                     >< MboI

```

FIGURE 13. 36

50/116

```

                >< RsaI          >< DpnII
                >< MaeII        >< DpnI      > < SspI
                >< Csp6I        >< BstYI     HinPII ><
    >< Tru9I
    >< RmaI          >< BsaAI        >< BspMI     Hin6I ><
    >< MaeI          >< AflIII       >< BspAI      HhaI ><
>< BspWI>< MseI          >< AfaI    >< AlwI>< Bsp143I   CfoI ><
AATGCTAGTT AAACAAGGAG ATGATTACGT GTACCTGCCT TACCCAGATC CATCAAGAAT ATTAGGCGCA
    15830      15840      15850      15860      15870      15880      15890

                >< RsaI          >< SfaNI
                >< Csp6I        >< MaeIII
                >< AfaI        BsrI ><
GGCTGTTTTG TCGATGATAT TGTCAAAACA GATGGTACAC TTATGATTGA AAGGTTCTGT TCACTGGCTA
    15900      15910      15920      15930      15940      15950      15960

    > < FokI
    >< BspWI
TTGATGCTTA CCCACTTACA AAACATCCTA ATCAGGAGTA TGCTGATGTC TTTCACCTGT ATTTACAATA
    15970      15980      15990      16000      16010      16020      16030

                >< Van91I
                >< PflMI
                >< NspI
                > < Pali>< NspHI
                > < MscI>< NlaIII
                > < HaeIII
                > < BsuRI
                >< BsrI
                >< EaeI >< BslI >< NspI
                > < BshI>< BsiYI >< NspHI
                >< NlaIII      >< AflIII >< AflIII
    >< MaeIII      >< AluI > < BalI>< AccB7I >< NlaIII
CATTAGAAAG TTACATGATG AGCTTACTGG CCACATGTTG GACATGTATT CCGTAATGCT AACTAATGAT
    16040      16050      16060      16070      16080      16090      16100

                >< RsaI> < NlaIV
                >< MnlI
                >< Csp6I      >< DdeI          >< RsaI
                >< BsrI >< MnlI      >< Csp6I
                >< AfaI> < BscBI      >< AfaI          SfcI ><
AACACCTCAC GGTACTGGGA ACCTGAGTTT TATGAGGCTA TGTACACACC ACATACAGTC TTGCAGGCTG
    16110      16120      16130      16140      16150      16160      16170

                >< NlaIV
                >< EcoNI
                >< Eco31I
                >< Eco64I>< BsmAI
                >< BscBI >< BslI
                >< BanI      >< BsiYI
                >< AciI      >< BsaI
                >< AccB1I>< Alw26I BbvI ><
>< BspWI
TAGGTGCTTG TGTATTGTGC AATTCACAGA CTTCACTTCG TTGCGGTGCC TGTATTAGGA GACCATTCCT
    16180      16190      16200      16210      16220      16230      16240

                >< Tth111I
                >< Fnu4HI      >< NlaIII
                >< BspWI >< AspI          > < Tru9I
                >< MseI
ATGTTGCAAG TGCTGCTATG ACCATGTCAT TTCAACATCA CACAAATTAG TGTGTCTGT TAATCCCTAT
    16250      16260      16270      16280      16290      16300      16310

                >< ScrFI
                >< MvaI

```

FIGURE 13.37

51/116

```

>< EcoRII
>< Ecl136I
>< DsaV
>< BstOI
>< BstNI
>< BsiLI
>< BsaJI
>< ApyI
GTTTGCAATG CCCAGGTTG TGATGTCAC TATGTGACAC AACTGTATCT AGGAGGTATG AGCTATTATT
16320 16330 16340 16350 16360 16370 16380

>< MaeIII >< MaeIII >< MaeIII >< MaeI >< AluI
GCAAGTCACA TAAGCCTCCC ATTAGTTTTC CATTATGTGC TAATGGTCAG GTTTTGGT TATACAAAAA
16390 16400 16410 16420 16430 16440 16450

>< NspI >< NspI
>< NspHI >< Tth111I >< NspHI
>< NlaIII>< MaeIII>< MaeIII >< NlaIII
>< AflIII >< AspI >< AflIII
CACATGTGTA GGCAGTGACA ATGTCAC TCAATGCG ATAGCAACAT GTGATTGGAC TAATGCTGGC
16460 16470 16480 16490 16500 16510 16520

>< RsaI
>< PleI
>< DdeI
>< Csp6I
>< BsmAI >< HinfI >< MnlI
>< Alw26I >< HindIII DdeI ><
>< AfaI >< AluI >< Fnu4HI >< BbvI
GATTACATAC TTGCAACAC TTGTACTGAG AGACTCAAGC TTTTCGCAGC AGAAACGCTC AAAGCCACTG
16530 16540 16550 16560 16570 16580 16590

>< ThaI
>< ScaI
>< RsaI >< RsaI
>< MvnI
>< Csp6I >< Csp6I
>< BstUI
>< Bsp50I
>< Tru9I
>< MseI >< NdeI >< AfaI >< AfaI
>< AluI >< AccII MnlI >
AGGAAACATT TAAGCTGTCA TATGGTATTG CCACTGTACG CGAAGTACTC TCTGACAGAG AATTGCATCT
16600 16610 16620 16630 16640 16650 16660

MaeIII ><
>< MaeIII
>< EcoO65I
>< Eco91I
>< BstPI
>< BstEII
>< BsrI
>< SfaNI >< RmaI
>< NlaIII >< MaeI
TTCATGGGAG GTTGGAAC CTAAGACC ACCATTGAACAGA AACTATGTCT TTACTGGTTA CCGTGTAAC
16670 16680 16690 16700 16710 16720 16730

RsaI ><
>< MnlI
>< HphI
>< RsaI >< RsaI
>< Csp6I >< Csp6I >< SfaNI Csp6I ><
>< AfaI >< AfaI >< MaeIII >< HphI AfaI ><
AAAAATAGTA AAGTACAGAT TGGAGAGTAC ACCTTTGAAA AAGGTGACTA TGGTGATGCT GTTGTGTACA
16740 16750 16760 16770 16780 16790 16800

```

FIGURE 13. 38

```

>< RsaI
>< Csp6I
>< AfaI
GAGGTACTAC GACATACAAG TTGAATGTTG GTGATTACTT TGTGTTGACA TCTCACACTG TAATGCCACT
16810      16820      16830      16840      16850      16860      16870

>< VneI
>< SnoI
  >< SduI
  >< NspII
  >< HgiAI
>< DraIII
  >< Bsp1286I
  >< BmyI
  >< ApaLI
  >< Alw44I
  >< Alw21I
  >< RmaI
  >< MaeI
  >< BspWI
  >< DraIII
  >< RsaI
  >< Csp6I
  >< BsrI
  >< AfaI
  >< DdeI
TAGTGCACCT ACTCTAGTGC CACAAGAGCA CTATGTGAGA ATTACTGGCT TGTACCCAAC ACTCAACATC
16880      16890      16900      16910      16920      16930      16940

  StyI ><
    SinI >
    Sau96I >
    NspIV >
    EcoT14I ><
    Eco47I >
    Eco130I ><
    >< ScaI Cfr13I >
    BssT1I ><
    >< SphI >< RsaI BsiZI >
    >< PaeI BsaJI ><
    >< NlaIII Bme18I >
    >< NspI>< Csp6I AvaII >
    >< NspHI>< AfaI AsuI >
TCAGATGAGT TTTCTAGCAA TGTGCAAAT TATCAAAAGG TCGGCATGCA AAAGTACTCT ACACTCCAAG
16950      16960      16970      16980      16990      17000      17010

  >< ScrFI
    >< RsaI
  >< MvaI
>< EcoRII
  >< Ecl136I
  >< Csp6I
  >< BstOI
  >< BstNI
>< XcmI >< BslI
>< NspHII >< BsiYI
  >< BsiLI
  >< ApyI >< BsrI
  >< DsaV>< AfaI >< HinfI>< PleI
GACCACCTGG TACTGGTAAG AGTCATTTTG CCATCGGACT TGCTCTCTAT TACCCATCTG CTCGCATAGT
17020      17030      17040      17050      17060      17070      17080

  >< SfaNI
    >< SphI >< PvuII
    >< PaeI >< Psp5I
    >< NspI >< NspBII
    >< NspHI >< Fnu4HI
  >< Bst1107I >< NlaIII>< BspWI >< Tru9I >< SspI
>< AccI >< NlaIII >< AluI >< BbvI >< MseI
GTATACGGCA TGCTCTCATG CAGCTGTTGA TGCCCTATGT GAAAAGGCAT TAAAATATTT GCCCATAGAT
17090      17100      17110      17120      17130      17140      17150

```

FIGURE 13.39

53/116

```

> < ThaI
>< ThaI
> < MvnI
>< MvnI >< ThaI
> < HinPII
>< HinPII
>< HinPII >< MvnI
> < Hin6I
>< Hin6I
> < HhaI
>< HhaI >< HhaI
> < CfoI
>< CfoI >< CfoI
> < BstUI
>< BstUI >< BstUI
>< BssHII
>< BspMI
> < Bsp50I
>< Bsp50I>< Bsp50I
RmaI >
MaeI >
>< TfiI >< Hin6I> < AccII
>< HinfI >< AccII >< AccII > < EcoRI
AAATGTAAGTA GAATCATACC TGCGCGTGCG CGCGTAGAGT GTTTTGATAA ATTCAAAGTG AATTCAACAC
17160 17170 17180 17190 17200 17210 17220

>< Zsp2I
>< Ppu10I
>< NsiI
>< Mph1103I
>< EcoT22I
>< BsgI > < AvaIII >< DrdI
TAGAACAGTA TGT TTTCTGC ACTGTAAATG CATTGCCAGA AACAACTGCT GACATTGTAG TCTTTGATGA
17230 17240 17250 17260 17270 17280 17290

>< RmaI
>< MaeI >< MaeII
AATCTCTATG GCTACTAATT ATGACTTGAG TGTGTGCAAT GCTAGACTTC GTGCAAAACA CTACGTCTAT
17300 17310 17320 17330 17340 17350 17360

>< Sau3AI
>< NdeII
>< MboI
>< DpnII
>< DpnI
>< BspAI
>< AlwI>< Bsp143I > < AciI >< RmaI
>< MaeI SspI ><
ATTGGCGATC CTGCTCAATT ACCAGCCCC CGCACATTGC TGAATAAGG CACACTAGAA CCAGAATATT
17370 17380 17390 17400 17410 17420 17430

>< SinI
>< Sau96I
>< NspIV >< StyI
>< NspHII >< NspI
>< Eco47I >< NspHI
>< Cfr13I >< NlaIII
>< BsiZI >< EcoT14I
>< BsgI >< Eco130I
>< Bme18I >< BssT1I
>< AvaII >< BsaJI
>< Tru9I
>< MseI
>< AsuI> < AflIII
TTAATTGAGT GTGCAGACTT ATGAAAACAA TAGGTCCAGA CATGTTCCCTT GGAACCTGTC GCCGTTGTCC
17440 17450 17460 17470 17480 17490 17500

```

FIGURE 13. 40

FIGURE 13. 41

FIGURE 13.42

```

TGTTCGACACT GAAAATAACA CAGAATTCAC CAGAGTTAAT GCAAAACCTC CACCAGGTGA CCAGTTTAAA
18350      18360      18370      18380      18390      18400      18410

      >< ScrFI
      >< MvaI
      >< EcoRII
      >< Ecl136I
      >< DsaV
      >< BstOI
      >< BstNI
      >< BsiLI
      >< BsaJI
      >< NlaIII
      >< ApyI
      >< RsaI
      >< DdeI ><
      >< Tru9I>< Csp6I
      >< MseI >< AfaI
CATCTTATAC CACTCATGTA TAAAGGCTTG CCCTGGAATG TAGTGCGTAT TAAGATAGTA CAAATGCTCA
18420      18430      18440      18450      18460      18470      18480

      >< NlaIII
      >< HinPII
      >< Tth111I
      >< Hin6I
      >< HinfI
      >< HhaI
      >< AspI >< PleI >< CfoI >< AluI
GTGATACT GAAAGGATTG TCAGACAGAG TCGTGTTCTG CCTTTGGGCG CATGGCTTTG AGCTTACATC
18490      18500      18510      18520      18530      18540      18550

      >< SinI
      >< Sau96I
      >< NspIV
      >< NspHII
      >< Eco47I
      >< Cfr13I
      >< ScaI
      >< BsiZI
      >< RsaI
      >< Bme18I
      >< Csp6I
      >< AvaII
      >< MaeII
      >< AfaI
      >< AsuI
      >< AflIII
      >< MaeIII>< MaeII
AATGAAGTAC TTTGTCAAGA TTGGACCTGA AAGAACGTGT TGTCTGTGTG ACAAACGTGC AACTTGCTTT
18560      18570      18580      18590      18600      18610      18620

      >< TfiI
      >< Tth111I
      >< HinfI
      >< AspI
TCTACTTCAT CAGATACTTA TGCCTGCTGG AATCATTCTG TGGGTTTTGA CTATGTCTAT AACCCATTTA
18630      18640      18650      18660      18670      18680      18690

      >< ScrFI
      >< RsaI ><
      >< MvaI
      >< EcoRII
      >< Ecl136I ><
      >< DsaV
      >< Csp6I ><
      >< BstXI ><
      >< MaeIII
      >< Eco651
      >< Eco91I
      >< BstPI
      >< ApyI
      >< Eco57I> >< BstEII
      >< MaeIII >< NlaIII
      >< AfaI ><
TGATTGATGT TCAGCAGTGG GGCTTTACGG GTAACCTTCA GAGTAACCAT GACCAACATT GCCAGGTACA
18700      18710      18720      18730      18740      18750      18760

      >< SfaNI
      >< RmaI
      >< NspI
      >< NspHI

```

FIGURE 13.43

```

                >< NlaIII                >< RmaI
                >< MaeI                >< NlaIII                Tru9I ><
>< NlaIII    >< BspWI                >< MaeI                >< NlaIII
    > < AflIII                >< BspHI                MseI ><
TGGAAATGCA CATGTGGCTA GTTGTGATGC TATCATGACT AGATGTTTAG CAGTCCATGA GTGCTTTGTT
    18770      18780      18790      18800      18810      18820      18830

    >< ThaI
    >< MvnI
>< HinPII
>< Hin6I
    >< HhaI
    >< CfoI
    >< BstUI                >< EcoNI> < MnlI
    >< Bsp50I                >< BslI                >< Tru9I
    >< AccII                >< BsiYI                >< DdeI >< MseI
AAGCGCGTTG ATTGGTCTGT TGAATACCTT ATTATAGGAG ATGAAGTCTG GGTAAATTCT GCTTGCAGAA
    18840      18850      18860      18870      18880      18890      18900

    >< RsaI
    >< Csp6I
    >< AfaI    >< NlaIII                >< BspWI                >< MboII                > < NlaIII
    >< AfaI    >< NlaIII                >< BspWI                >< BsrI    >< BspHI
AAGTACAACA CATGGTTGTG AAGTCTGCAT TGCTTGCTGA TAAGTTTCCA GTTCTTCATG ACATTGGAAA
    18910      18920      18930      18940      18950      18960      18970

                >< SauI
                >< MstII
                >< Eco81I
                >< DdeI                NlaIII ><
                >< CvnI                >< EspI
                >< Bsu36I                >< Eco57I MaeIII ><
                >< Bse21I                >< DdeI
                >< AxyI                >< CelII
                >< AocI    >< MnlI    >< SfaNI                >< Bpu1102I
TCCAAAGGCT ATCAAGTGTG TGCCTCAGGC TGAAGTAGAA TGGAAGTTCT ACGATGCTCA GCCATGTAGT
    18980      18990      19000      19010      19020      19030      19040

                >< MnlI                >< Ksp632I
                >< HindIII                >< EarI
                >< AluI    >< MboII                >< Eam1104I
GACAAAGCTT ACAAATAGA GGAAGTCTTC TATTCTTATG CTACACATCA CGATAAATTC ACTGATGGTG
    19050      19060      19070      19080      19090      19100      19110

                >< Sau3AI
                >< NdeII
                >< MboI
                >< MaeII> < MaeIII
                >< DpnII
                >< DpnI
                >< BspAI                HinfI >
                >< MaeIII >< Bsp143I                >< MunI                DrdI ><
TTTGTGTTGTT TTGAATTGT AACGTTGATC GTTACCCAGC CAATGCAATT GTGTGTAGGT TTGACACAAG
    19120      19130      19140      19150      19160      19170      19180

                >< ScrFI                Zsp2I ><
                >< MvaI                >< SphI
                >< EcoRII                > < Ppu10I
                >< EcoRII                >< PaeI
                >< EcoRII                >< NspI
                >< EcoRII                >< NspHI
                >< EcoRII                >< NlaIII
                >< EcoRII                Mph1103I ><

```

FIGURE 1344

58/116.

```

                >< Ecl136I
            >< DsaV
                >< BstOI
                >< BstNI
                >< BsiLI
                >< ApyI
    >< PleI
AGTCTTGTCA AACTTGAAC TACCAGGCTG TGATGGTGGT AGTTTGTATG TGAATAAGCA TGCATTCCAC
  19190      19200      19210      19220      19230      19240      19250

                >< Tru9I
                > < MunI
            >< TthHB8I
            >< BcgI/a >< TaqI
            >< AluI
            >< MseI
            >< DraI
            >< BcgI
ACTCCAGCTT TCGATAAAAG TGCATTTACT AATTTAAAGC AATTGCCTTT CTTTTACTAT TCTGATAGTC
  19260      19270      19280      19290      19300      19310      19320

                >< PleI
                >< NlaIII
                >< BsmAI
            >< HinfI>< Alw26I
            SfaNI ><
                >< MaeII
                BsaAI ><
                AflIII ><
CTTGTGAGTC TCATGGCAAA CAAGTAGTGT CGGATATTGA TTATGTTCCA CTCAAATCTG CTACGTGTAT
  19330      19340      19350      19360      19370      19380      19390

                Zsp2I >
                >< ScaI
                Ppu10I ><
                >< RsaINsiI >
                Mph1103I >
                >< SfaNEcoT22I >
                > < RsaI >< Csp6I
                >< Csp6I
                >< NlaIII> < AfaI >< AfaI
                AvaIII ><
TACACGATGC AATTTAGGTG GTGCTGTTTG CAGACACCAT GCAAATGAGT ACCGACAGTA CTTGGATGCA
  19400      19410      19420      19430      19440      19450      19460

            >< FokI
TATAATATGA TGATTTCTGC TGGATTTAGC CTATGGATTT ACAAACAATT TGATACTTAT AACCTGTGGA
  19470      19480      19490      19500      19510      19520      19530

            >< ScrFI
            >< MvaI
            >< MaeIII
            >< EcoRII
            >< Ecl136I
            >< DsaV
            >< BstOI
            >< BstNI
            >< BsiLI
            >< ApyI
            >< Tru9I
            >< MseI
ATACATTTAC CAGGTTACAG AGTTTAGAAA ATGTGGCTTA TAATGTTGTT AATAAAGGAC ACTTTGATGG
  19540      19550      19560      19570      19580      19590      19600

            >< SgrAI
            >< NaeI
            >< MspI
            >< HpaII
            >< HapII
            >< Cfr10I
            >< BspWI
            > < VspI
            > < Tru9I
            > < MseI
            > < AsnI
            > < AseI
ACACGCCGGC GAAGCACCTG TTTCCATCAT TAATAATGCT GTTTACACAA AGGTAGATGG TATTGATGTG
  19610      19620      19630      19640      19650      19660      19670

```

FIGURE 13. 45

```

>< XhoII
>< Sau3AI
>< NdeII
>< MflI
>< MboI
>< DpnII
  >< DpnI
>< BstYI
>< BspAI
  >< Bsp143I
>< BglII
GAGATCTTTG AAAATAAGAC AACACTTCCT GTTAATGTTG CATTTGAGCT TTGGGCTAAG CGTAACATTA
  19680      19690      19700      19710      19720      19730      19740

                                >< MaeIII
                                >< EspI
                                >< DdeITru9I ><
                                >< CeliIMseI ><
                                >< Bpu1102I
                                >< AluI
                                >< Tru9I
                                >< MseI
                                >< Fnu4HI
                                >< EcoRV
                                >< Eco32I
                                >< BbvI
                                >< BsrI
                                >< MseI
                                >< Tru9I
AACCAGTGCC AGAGATTAAG ATACTCAATA ATTTGGGTGT TGATATCGCT GCTAATACTG TAATCTGGGA
  19750      19760      19770      19780      19790      19800      19810

                                >< NspI
                                >< NspHI
                                >< NlaIII
                                >< BsgI
                                >< AflIII
CTACAAAAGA GAAGCCCCAG CACATGTATC TACAATAGGT GTCTGCACAA TGACTGACAT TGCCAAGAAA
  19820      19830      19840      19850      19860      19870      19880

                                >< DdeI>< MboII
                                >< AccI
CCTACTGAGA GTGCTTGTTT TTTCACTTACT GTCTTGTTTG ATGGTAGAGT GGAAGGACAG GTAGACCTTT
  19890      19900      19910      19920      19930      19940      19950

                                SinI ><
                                Sau96I ><
                                NspIV ><
                                NspHII ><
                                NlaIV ><
                                Eco47I ><
                                Cfr13I ><
                                >< BslI
                                BsiZI ><
                                >< BsiYI
                                BscBI ><
                                Bme18I ><
                                AvaII ><
                                AsuI ><
                                >< Tru9I
                                >< MseI
TTAGAAACGC CCGTAATGGT GTTTTAATAA CAGAAGGTTT AGTCAAAGGT CTAACACCTT CAAAGGGACC
  19960      19970      19980      19990      20000      20010      20020

                                >< VspI
                                >< Tru9I
                                >< PleI
                                >< MseI
                                >< RmaI
                                >< NheI
                                >< MaeI
                                >< HgaI>< AluI
                                >< HinfI>< AseI
                                >< TfiI
                                >< HinfI
                                >< MseI
                                >< Tru9I
                                >< Tru9I
                                >< MseI
AGCACAAGCT AGCGTCAATG GAGTCACATT AATTGGAGAA TCAGTAAAAA CACAGTTTAA CTACTTTAAG
  20030      20040      20050      20060      20070      20080      20090

                                >< DdeI
                                >< MnlI
                                >< Tru9I ><
                                >< BsmAI
                                >< DdeI

```

FIGURE 1346

```

>< AccI                                     >< Alw26I >< BfrIMseI ><
AAAGTAGACG GCATTATTCA ACAGTTGCCT GAAACCTACT TTACTCAGAG CAGAGACTTA GAGGATTTTA
20100      20110      20120      20130      20140      20150      20160

>< TthHB8I
>< TaqI
>< SstI
>< SduI
>< SacI
>< PaeR7I
>< NspIII
>< NspII
>< HgiAI
>< Eco88I
>< XhoI>< Eco24I
>< Ecl136II
>< SlaI>< Bsp1286I
>< CcrI>< BmyI
>< BcoI>< BanII
>< Ama87I
>< AvaI>< Alw21I
>< XcmI
>< Sau3AI
>< NdeII
>< MboI
>< DpnII
>< DpnI
>< BspAI
>< Bsp143I
>< AluI
>< EcoRI
>< FokIAluI ><
AGCCCAGATC ACAAATGGAA ACTGACTTTC TCGAGCTCGC TATGGATGAA TTCATACAGC GATATAAGCT
20170      20180      20190      20200      20210      20220      20230

>< TthHB8I
>< TaqI
>< SfuI
>< NspV
>< LspI
>< Csp45I
>< BstBI
>< Bsp119I
>< BsiCI
>< Bpu14I
>< AsuII
>< BcgI
>< MboII
>< BbsI
>< Tru9I ><
>< NlaIII >< AciIMseI ><
CGAGGGCTAT GCCTTCGAAC ACATCGTTTA TGGAGATTTC AGTCATGGAC AACTTGGCGG TCTTCATTTA
20240      20250      20260      20270      20280      20290      20300

>< HphI
>< HinPII
>< Hin6I
>< EspI
>< DdeI
>< CelII
>< Bpu1102I
>< BfrI
>< HhaI >< TfiI
>< HaeII
>< Eco47III
>< CfoI >< HinfI
>< Bsp143II
>< Tru9I
>< MseI
>< MnlI
ATGATAGGCT TAGCCAAGCG CTCACAAGAT TCACCACTTA AATTAGAGGA TTTTATCCCT ATGGACAGCA
20310      20320      20330      20340      20350      20360      20370

>< MstI
>< HinPII
>< Hin6I
>< HhaI
>< FspI
>< FdiII
>< CfoI
>< SfaNI
>< AviII
>< Sau3AI ><
>< NdeII ><
>< MboI ><
>< DpnII ><
>< DpnI ><
>< BspAI ><
>< Bsp143I ><
CAGTGAAAAA TTACTTCATA ACAGATGCGC AAACAGGTTC ATCAAAATGT GTGTGTTCTG TGATTGATCT
20380      20390      20400      20410      20420      20430      20440

>< TthHB8I

```

FIGURE 13.47

61/116

```

    >< Tth111I
      >< TaqI
    >< AspI          > < MaeIII          MaeIII ><
TTTACTTGAT GACTTTGTCG AGATAATAAA GTCACAAGAT TTGTCAGTGA TTTCAAAAGT GGTCAAGGTT
    20450      20460      20470      20480      20490      20500      20510

                                >< NspI
                                >< NspHI
                                >< NlaIII
                                >< FokI

    >< MunI          > < NlaIII          >< AflIII
ACAATTGACT ATGCTGAAAT TTCATTTCATG CTTTGGTGTA AGGATGGACA TGTTGAAACC TTCTACCCAA
    20520      20530      20540      20550      20560      20570      20580

                                >< SfaNI
                                >< ScrFI
                                >< MvaI
                                >< EcoRII
                                >< Ecl136I
                                >< DsaV
                                >< BstOI          >< SfaNI
                                >< BstNI          >< RsaI   BspWI ><
                                >< BsiLI          > < Csp6I   BsmI >
                                >< ApyI          >< AfaI   BscCI ><
    >< BspWI
AACTACAAGC AAGTCAAGCG TGGCAACCAG GTGTTGCGAT GCCTAACTTG TACAAGATGC AAAGAATGCT
    20590      20600      20610      20620      20630      20640      20650

    >< Eco57I >< MaeIII          >< HphI
TCTTGAAAAG TGTGACCTTC AGAATTATGG TGAAAATGCT GTTATACCAA AAGGAATAAT GATGAATGTC
    20660      20670      20680      20690      20700      20710      20720

                                > < RsaI
                                >< Csp6I

    >< Bst1107I          >< Tru9I          >< AluI
    >< AccI          >< MseI          > < AfaINlaIII ><
GCAAAGTATA CTCAACTGTG TCAATACTTA AATACACTTA CTTTAGCTGT ACCCTACAAC ATGAGAGTTA
    20730      20740      20750      20760      20770      20780      20790

                                >< ScrFI
                                >< RsaI
                                >< MvaI
    >< EcoRII >< NspBII
    >< Ecl136I          >< SduI
    > < Csp6I          >< NspII
    >< BstOI >< PvuII>< HgiAI
    >< BstNI          >< DdeI
    >< BsiLI >< Psp5I>< Bsp1286I
    >< ApyI >< AluI >< BmyI
    >< DsaV>< AfaI          >< Alw21I
TTCAC TTTGG TGCTGGCTCT GATAAAGGAG TTGCACCAGG TACAGCTGTG CTCAGACAAT GGT TGCCAAC
    20800      20810      20820      20830      20840      20850      20860

                                >< XhoII
                                >< Tru9I
                                >< Sau3AI
                                >< NdeII
    >< TthHB8I >< MseI
    >< MflI
    >< MboI
    >< MamI
    >< DpnII
    >< TfiI >< DpnI

```

FIGURE 13. 48

```

                >< BstYI                > < TfiI
                >< BspAI                > < HinfI
                >< HinfI>< Bsp143I        >< Esp3I                >< Tru9I
                >< BsiBI                >< Tth111I        >< BsmBI        >< MseI
                >< BsaBI                >< BsmAI                > < BsmAI
    >< BsrI        >< TaqI >< BglII        >< AspI        >< Alw26I >< HgaI> < Alw26I
TGGCACACTA CTTGTCGATT CAGATCTTAA TGACTTCGTC TCCGACGCAG ATTCTACTTT AATTGGAGAC
    20870        20880        20890        20900        20910        20920        20930

                >< StyI
                >< SinI
                >< Sau96I
                >< RmaI
                >< NspIV
                >< NspHII ><
                >< MaeI
                >< EcoT14I
                >< Eco47I
                >< Eco130I
                >< Cfr13I
                >< BssT1I
                >< Bsi2I
                >< BsaJI
                >< Bml18I
                >< BlnI
                >< AvrII
                >< AvaII
                >< AsuI
                AflIII ><
    >< RsaI                >< Bml18I
    > < Csp6I                >< AvaII
    >< AfaI                >< AsuI
TGTGCAACAG TACATACGGC TAATAAATGG GACCTTATTA TTAGCGATAT GTATGACCCT AGGACCAAAC
    20940        20950        20960        20970        20980        20990        21000

    >< NspI
    >< NspHI
    >< NlaIII >< PleI
    >< MaeIII                >< HinfI
    ATGTGACAAA AGAGAATGAC TCTAAAGAAG GGTTTTTCAC TTATCTGTGT GGATTATATA AGCAAAAAC
    21010        21020        21030        21040        21050        21060        21070

    >< ScrFI
    >< MvaI
    >< EcoRII
    >< Ecl136I
    >< DsaV
    >< BstOI
    >< BstNI
    >< BsiLI
    >< BsaJI
    >< BsaJI        >< SfcI                >< BsmI                >< BsmI        AsuI >
    >< ApyI        > < AluI                >< BscCI                >< BscCIHindIII ><< AluI
AGCCCTGGGT GGTTCATAG CTGTAAAGAT AACAGAGCAT TCTTGAATG CTGACCTTTA CAAGCTTATG
    21080        21090        21100        21110        21120        21130        21140

                >< Zsp2I
                >< Ppu10I
                >< NsiI
                >< Mph1103I        Tru9I ><
                >< EcoT22I                >< MseI
    >< BshI        >< NlaIII>< AluI        >< BcgI        >< AvaIII >< SfaNIBcgI/a ><
GGCCATTTCT CATGGTGGAC AGCTTTTGTG ACAAATGTAA ATGCATCATC ATCGGAAGCA TTTTAAATTG
    21150        21160        21170        21180        21190        21200        21210

```

FIGURE 13.49

```

                                >< Zsp2I
                                >< SphI
                                >< Ppu10I
                                >< PaeI
                                >< NspI
                                >< NspHI
                                >< NsiI
                                >< NlaIII
                                > < NlaIII
                                >< Mph1103I
                                >< EcoT22I
                                > < AvaIII >< MnlI
GGGCTAACTA TCTTGGCAAG CCGAAGGAAC AAATTGATGG CTATACCATG CATGCTAACT ACATTTTCTG
    21220      21230      21240      21250      21260      21270      21280

                                >< MboII
                                >< GsuI
                                >< BsrI
                                >< BpmI
                                >< BbsI
                                >< NlaIII >< MnlI
GAGGAACACA AATCCTATCC AGTTGTCTTC CTATTCACCTC TTTGACATGA GCAAATTTCC TCTTAAATTA
    21290      21300      21310      21320      21330      21340      21350

                                >< Tru9I
                                >< MseI
                                >< Esp4I> < TfiI
                                >< BsmAI
                                >< Alw26I
                                >< AflII> < HinfI
                                >< Ksp632I ><
                                >< MboII >< EarI
                                >< Eam1104I ><
AGAGGAACTG CTGTAATGTC TCTTAAGGAG AATCAAATCA ATGATATGAT TTATTCTCTT CTGGAAAAAG
    21360      21370      21380      21390      21400      21410      21420

                                >< Tru9I
                                >< MseI
                                >< HindII
                                >< HincII
                                >< HpaI AflIII >
GTAGGCTTAT CATTAGAGAA AACAACAGAG TTGTGGTTTC AAGTGATATT CTTGTTAACA ACTAAACGAA
    21430      21440      21450      21460      21470      21480      21490

                                >< VneI
                                >< SnaI
                                >< SduI
                                >< NspII
                                >< HpaII
                                >< HgiAI
                                >< HapII
                                >< Cfr10I
                                >< Bsp1286I
                                >< MspI>< BmyI
                                >< ApaI
                                >< Alw44I
                                >< AgeI >< Alw21I
CATGTTTATT TTCTTATTAT TTCTTACTCT CACTAGTGGT AGTGACCTTG ACCGGTGCAC CACTTTTGAT
    21500      21510      21520      21530      21540      21550      21560

                                > < AluI
                                >< MnlI
GATGTTCAAG CTCCTAATTA CACTCAACAT ACTTCATCTA TGAGGGGGGT TTACTATCCT GATGAAATTT
    21570      21580      21590      21600      21610      21620      21630

                                >< Sau3AI

```

FIGURE 13. 50

```

>< NdeII
>< MboI
>< DpnII
  >< DpnI          >< Tru9I
>< BspAI          >< MseI > < MboII
  >< Bsp143I       >< DdeI          >< MaeIII
TTAGATCAGA CACTCTTTAT TTAACCTCAGG ATTTATTTCT TCCATTTTAT TCTAATGTTA CAGGGTTTCA
  21640      21650      21660      21670      21680      21690      21700

  >< VspI
  >< Tru9I
  >< MseI
  >< AsnI          >< Tru9I          >< FokI
  >< AseI >< MaeII >< MseI >< BbvI          > < Fnu4HI
TACTATTAAT CATACGTTTG GCAACCCTGT CATACCTTTT AAGGATGGTA TTTATTTTGC TGCCACAGAG
  21710      21720      21730      21740      21750      21760      21770

          >< BslI
          >< DsaI>< BsiYI          >< NlaIII
          >< BsaJI          > < MaeIII
AAATCAAATG TTGTCCGTGG TTGGGTTTTT GGTTCACCA TGAACAACAA GTCACAGTCG GTGATTATTA
  21780      21790      21800      21810      21820      21830      21840

          >< NspI
>< Tru9I          >< NspHI
>< MseI          >< NlaIII
  >< HphI          >< MaeIII          >< MaeIII
TTAACAATTC TACTAATGTT GTTATACGAG CATGTAACCT TGAATTGTGT GACAACCCTT TCTTTGCTGT
  21850      21860      21870      21880      21890      21900      21910

  >< StyI          >< Zsp2I
    >< NlaIII          >< Tru9I
  >< NcoI >< RsaI          >< Ppu10I TthHB8I ><
  >< EcoT14I          >< NsiI          >< TaqI
  >< Eco130I          >< MseI          SfaNI ><
  >< DsaI>< Csp6I          >< Mph1103I RsaI ><
  >< BssT1I          >< TthHB8I >< EcoT22I Csp6I ><
  >< BsaJI>< AfaI          >< TaqI >< AfaIII AfaI ><
TTCTAAACCC ATGGGTACAC AGACACATAC TATGATATTC GATAATGCAT TTAATTGCAC TTTCGAGTAC
  21920      21930      21940      21950      21960      21970      21980

          >< Tru9I
          >< MseI
          >< DraI
ATATCTGATG CCTTTTCGCT TGATGTTTCA GAAAAGTCAG GTAATTTTAA ACACCTTACGA GAGTTTGTGT
  21990      22000      22010      22020      22030      22040      22050

          >< Sau3AI
          >< NdeII
          >< MboI
          >< DpnII
          >< DpnI
          >< BspAI
>< Tru9I
>< MseI
>< DraI          >< SfcI Bsp143I ><
TTAAAAATAA AGATGGGTTT CTCTATGTTT ATAAGGGCTA TCAACCTATA GATGTAGTTC GTGATCTACC
  22060      22070      22080      22090      22100      22110      22120

          >< Tru9I
  >< Tru9I          > < Tru9I          >< MseI
  >< MseI          > < MseI          >< MnlI
TTCTGGTTTT AACACTTTGA AACCTATTTT TAAGTTGCCT CTTGGTATTA ACATTACAAA TTTTAGAGCC
  22130      22140      22150      22160      22170      22180      22190

```

FIGURE 13.51

65/116

```

> < SduI>< SfcI
    >< PvuII
    >< Psp5I
> < NspII
    >< NspBII
> < MaeII > < Fnu4HI
> < Bsp1286I >< PstI      Tru9I >
    >< BmyI>< Fnu4HI      MseI >
    >< BspMI
    >< BbvI
    >< AluI
    >< BbvI
    >< HphI
ATTCTTACAG CCTTTTCACC TGCTCAAGAC ATTTGGGGCA CGTCAGCTGC AGCCTATTTT GTTGGCTATT
22200      22210      22220      22230      22240      22250      22260

    >< SfaNI
    >< RsaI
> < Csp6I
    >< AfaI
    >< AlwNI
    >< DraI
TAAAGCCAAC TACATTTATG CTCAAGTATG ATGAAAATGG TACAATCACA GATGCTGTTG ATTGTTCTCA
22270      22280      22290      22300      22310      22320      22330

    > < Tru9I
    > < MseI
    >< AluI
AAATCCACTT GCTGAACTCA AATGCTCTGT TAAGAGCTTT GAGATTGACA AAGGAATTTA CCAGACCTCT
22340      22350      22360      22370      22380      22390      22400

    >< SauI
    >< MstII
    >< Eco81I
    >< DdeI
    >< CvnI
    >< Bsu36I
    >< Bse21I
    >< AxyI
    >< TfiI
    >< MnlI
    >< AocI
    >< MnlI
    >< HinfI
    >< SspI
    >< MnlI
AATTCAGGG TTGTTCCCTC AGGAGATGTT GTGAGATTCC CTAATATTAC AAACCTGTGT CTTTTGGAG
22410      22420      22430      22440      22450      22460      22470

    >< Zsp2I
    >< Ppu10I
    >< NsiI
    > < NlaIII
    >< Mph1103I
    >< EcoT22I
    >< Tru9I
    >< MseI
    >< AvaIII
AGGTTTTTAA TGCTACTAAA TTCCCTTCTG TCTATGCATG GGAGAGAAAA AAAATTTCTA ATTGTGTTGC
22480      22490      22500      22510      22520      22530      22540

    >< SduI
    >< NspII
    >< HgiAI
    >< Bsp1286I
    >< BmyI
    >< Alw21I
    >< Tru9I
    >< MseI
    >< DdeI
TGATTACTCT GTGCTCTACA ACTCAACATT TTTTTC AACC TTTAAGTGCT ATGGCGTTTC TGCCACTAAG
22550      22560      22570      22580      22590      22600      22610

    >< Sau3AI
    >< NdeII
    >< MboI
    >< DpnII
    >< DpnI

```

FIGURE 13.52

66/116

```

    >< BspAI
    >< Bsp143I
TTGAATGATC TTTGCTTCTC CAATGTCTAT GCAGATTCTT TTGTAGTCAA GGGAGATGAT GTAAGACAAA
22620      22630      22640      22650      22660      22670      22680

    >< ScrFI
    >< MvaI
>< HinPII
>< Hin6I
    >< HhaI
    >< HaeII
    >< EcoRII
    >< Ecl136I
    >< DsaV
    >< CfoI
    >< BstOI
    >< BstNI
    >< Bsp143II
    >< BsiLI
    >< ApyI      > < BsrI
TAGCGCCAGG ACAAACTGGT GTTATTGCTG ATTATAATTA TAAATTGCCA GATGATTTCA TGGGTTGTGT
22690      22700      22710      22720      22730      22740      22750

    >< SfaNI
    >< RmaI
    >< MaeI
    >< BsrI
    >< NlaIII
CCTTGCTTGG AATACTAGGA ACATTGATGC TACTTCAACT GGTAATTATA ATTATAAATA TAGGTATCTT
22760      22770      22780      22790      22800      22810      22820

    >< Sau96I
    >< Pali
    >< NspIV
> < HindIII
    >< HaeIII
    >< EcoO109I
    >< DraII
    >< DdeI
    >< Cfr13I
    >< BsuRI
    >< BsiZI
    >< BshI
    >< BfrI >< PssI
    >< NlaIII >< AsuI>< BsmAI
    >< AluI      >< Alw26I
AGACATGGCA AGCTTAGGCC CTTTGAGAGA GACATATCTA ATGTGCCTTT CTCCCCTGAT GGCAAACCTT
22830      22840      22850      22860      22870      22880      22890

    >< Tru9I
    >< Pali
    >< MscI
    >< HaeIII
    >< EaeI>< MseI
    >< Tru9I      >< BsuRI
    >< MseI      >< BshI
    >< BspMI      >< BalI
    >< BsrI ><
GCACCCACC TGCTCTTAAT TGTATTGGC CATTAAATGA TTATGGTTTT TACACCACTA CTGGCATTGG
22900      22910      22920      22930      22940      22950      22960

    >< Sau96I ><
    >< PalINspIV ><
> < MspI NspHII ><
    >< HaeIII

```

FIGURE 13.53

```
> < HpaII Eco47I ><
    >< DsaI
> < HapII Cfr13I ><
    >< BsuRISinI ><
    >< GdiII BsiZI ><
        >< BsaJI
```

```
>< Tru9I    >< EaeI Bme18I ><
>< MseI >< Cfr10I    AvaII ><
>< DraI      >< BshI AsuI ><
```

```

      << Tru9I          << RsaI
<< Tru9I              << Csp6I
      << P1eI          BsrI <<
      << MseI          << BsrI
<< MseI              << AfaI
      << HinfI

```

```
>< Pali
>< HaeIII
>< GdiII
>< EaeI
>< BsuRI          TfiI ><
>< BshI           HinfI ><
```

```
> < XhoII
>< TthHB8I
>< TaqI
> < Sau3AI
> < NdeII
> < MflI
> < MboI
> < DpnII
    >< DpnI
> < BstYI
> < BspAI
```

```
>< ScrFI
>< MvaI
>< EcoRII
>< Ecl136I
>< DsaV
>< BstOI
>< BstNI
>< BsiLI
>< ApyI
```

>< HinP1I

FIGURE 13. 54

FIGURE 13.55

69/116

```

>< RsaI
>< Csp6I
>< AfaI
GTACAAAACC CCAACTTTGA AATATTTTGG TGGTTTAAAT TTTTCACAAA TATTACCTGA CCCTCTAAAG
  23810      23820      23830      23840      23850      23860      23870

>< MnlI
>< MnlI
>< DdeI >< MnlI
CCAATAAGA GGTCTTTTAT TGAGGACTTG CTCTTTAATA AGGTGACACT CGCTGATGCT GGCTTCATGA
  23880      23890      23900      23910      23920      23930      23940

>< XhoII
>< Sau3AI
>< StyI
>< RmaI
>< MaeI
>< EcoT14I
>< Eco130I
>< BssT1I
>< BsmI
>< BscCI
>< BsaJI
>< BlnI
>< AvrII
>< RmaI
>< NdeII
>< MflI
>< MboI
>< MaeI
>< VspI
>< HphI> < DpnI
>< Tru9I
>< MseI
>< AsnI
>< AseI
>< XhoII
>< Sau3AI
>< NdeII
>< MflI
>< MboI
>< MaeI
>< VspI
>< HphI> < DpnI
>< BstYI
>< BspAI
>< Bsp143I
>< BglII
>< MstI
>< HinPII
>< Hin6I
>< HhaI
>< FspI
>< FdiII
>< CfoI
>< AviII
AGCAATATGG CGAATGCCTA GGTGATATTA ATGCTAGAGA TCTCATTTGT GCGCAGAAGT TCAATGGACT
  23950      23960      23970      23980      23990      24000      24010

>< RmaIRsaI ><
>< MnlI
>< Fnu4HI
>< Fnu4HI Csp6I ><
>< BspWI >< BbvI
>< BbvI >< BspWI >< MaeIAfaI ><
TACAGTGTG CCACCTCTGC TCACTGATGA TATGATTGCT GCCTACACTG CTGCTCTAGT TAGTGGTACT
  24020      24030      24040      24050      24060      24070      24080

>< MboII
>< HinPII
>< Hin6I
>< HhaI
>< HaeII
>< Fnu4HI >< Ksp632I
>< CfoI >< EarI
>< FokI >< BspWI >< Eam1104I
>< BbvI
>< Bsp143II
GCCACTGCTG GATGGACATT TGGTGCTGGC GCTGCTCTTC AAATACCTTT TGCTATGCAA ATGGCATATA
  24090      24100      24110      24120      24130      24140      24150

Tru9I ><
MseI ><
>< MaeIII
GGTTCAATGG CATTGGAGTT ACCCAAAATG TTCTCTATGA GAACCAAAAA CAAATCGCCA ACCAATTTAA
  24160      24170      24180      24190      24200      24210      24220

MaeII ><
>< TfiI
>< HinfI
>< BbvI
>< AluI
CAAGGCGATT AGTCAAATTC AAGAATCACT TACAACAACA TCAACTGCAT TGGGCAAGCT GCAAGACGTT
  24230      24240      24250      24260      24270      24280      24290

>< Tru9I
>< MseI
>< HpaI
>< HindII
>< HincII>
>< BsmI
>< BscCI
>< Tru9I
>< MseI
>< DdeI
>< BfrI
>< AluI

```

FIGURE 13. 56

```

GTTAACCAGA ATGCTCAAGC ATTAAACACA CTTGTAAAC AACTTAGCTC TAATTTTGGT GCAATTTCAA
 24300      24310      24320      24330      24340      24350      24360

      >< ThaI
      >< SpoI
      >< NruI
      >< MvnI
      >< BstUI      >< TthHB8I
      >< Bsp68I      >< TaqI      >< RsaI
      >< EcoRV      >< Bsp50I      >< MnlI      >< Csp6I      >< Tru9I
      >< Eco32I      >< AccII      >< MnlI      >< AciI      >< AfaI      >< MseI
GTGTGCTAAA TGATATCCTT TCGCGACTTG ATAAAGTCGA GCGGAGGTA CAAATTGACA GGTAAATTAC
 24370      24380      24390      24400      24410      24420      24430

      >< MaeIII      >< BbvI      >< Fnu4HI      BbvI      ><
AGGCAGACTT CAAAGCCTTC AAACCTATGT AACACAACAA CTAATCAGGG CTGCTGAAAT CAGGGCTTCT
 24440      24450      24460      24470      24480      24490      24500

      >< Fnu4HI
      >< BspWI      >< DdeI      >< HindII
      >< BspWI      >< DdeI      >< HincII
GCTAATCTTG CTGCTACTAA AATGTCTGAG TGTGTTCTTG GACAATCAAA AAGAGTTGAC TTTTGTGGAA
 24510      24520      24530      24540      24550      24560      24570

      > < NspI
      > < NspHI
      > < NlaIII
      >< MaeIII
      >< NlaIII      >< MaeII
      >< MboII      >< FokI
      >< Fnu4HI      >< BbsI      BsaAI      ><
      >< AciI      >< BbvI      >< AflIII
AGGGCTACCA CTTATGTCC TTCCCACAAG CAGCCCCGCA TGGTGTGTC TTCCTACATG TCACGTATGT
 24580      24590      24600      24610      24620      24630      24640

      >< ScrFI
      >< MvaI
      >< EcoRII
      >< Ecl136I
      >< BstOI
      >< BstNI      >< HinPII
      >< MnlI      >< BslI      >< Hin6I
      >< DsaV      >< BsiYI      >< HhaI
      >< BsiLI      >< HaeII
      >< BsaJI      >< HphI      >< CfoI      >< NlaIII
      >< ApyI      >< Bsp143II      >< BspHI      EcoNI      ><
GCCATCCCAG GAGAGGAAGT TCACCACAGC GCCAGCAATT TGTCATGAAG GCAAAGCATA CTTCCCTCGT
 24650      24660      24670      24680      24690      24700      24710

      >< MnlI
      >< BslI      >< Tru9I
      >< BsiYI      >< MseI      >< MnlI
GAAGGTGTTT TTGTGTTTAA TGGCACTTCT TGGTTTATTA CACAGAGGAA CTTCTTTTCT CCACAAATAA
 24720      24730      24740      24750      24760      24770      24780

      >< DdeI      >< Tru9I
      >< BsmAI      >< SfaNI
      >< SfcI      >< Alw26I      >< MseI      >< AlwI      ><
TTACTACAGA CAATACATTT GTCTCAGGAA ATTGTGATGT CGTTATTGGC ATCATTAACA ACACAGTTTA
 24790      24800      24810      24820      24830      24840      24850

      >< Sau3AI
      >< NdeII

```

FIGURE 13.57

71/116

```

>< MboI          >< P1eI          > < ScaI
>< DpnII         >< MnlI          > < Ksp632I      > < RsaI
  >< DpnI         >< DdeI   >< HinfI      >< MboII
>< BspAI         >< BspWI          > < Eam1104I      >< Csp6I
  >< Bsp143I      >< AluI          > < EarI   > < AluI   > < AfaI   > < HphI
TGATCCTCTG CAACCTGAGC TTGACTCATT CAAAGAAGAG CTGGACAAGT ACTTCAAAAA TCATACATCA
  24860      24870      24880      24890      24900      24910      24920

  >< Sau3AI
  >< NdeII
  >< MboI
>< MamI
  >< DpnII
    >< DpnI
  >< BspAI
    >< Bsp143I
  >< BsiBI          >< Tru9I          >< HindII
  >< BsaBI          >< MseI           >< HincII          AciI ><
CCAGATGTTG ATCTTGCGCA CATTTCAGGC ATTAACGCTT CTGTCGTCAG CATTCAAAAA GAAATTGACC
  24930      24940      24950      24960      24970      24980      24990

          >< Tru9I
          > < TfiI
          >< MnlI          >< SmaI
  >< EcoNI          >< MseI
    >< BslI          > < HinfI
>< MnlI>< BsiYI      >< DraI
GCCTCAATGA GGTCGCTAAA AATTTAAATG AATCACTCAT TGACCTTCAA GAATTGGGAA AATATGAGCA
  25000      25010      25020      25030      25040      25050      25060

  >< StyI
  >< P1eI
  >< HaeIII
  >< EcoT14I
  >< Eco130I
  >< BsuRI
  >< BssT1I          NlaIII ><
  >< Tru9I>< BshI          MaeIII ><
  >< MseI   >< BsaJI          >< BstXI
ATATATTAAA TGGCCTTGGT ATGTTTGGCT CGGCTTCATT GCTGGACTAA TTGCCATCGT CATGGTTACA
  25070      25080      25090      25100      25110      25120      25130

          > < SphI
          > < PaeI
          > < NspI
          > < NspHI
          > < NlaIII
          >< MnlI>< BbvI   Fnu4HI ><
ATCTTGCTTT GTTGCATGAC TAGTTGTTGC AGTTGCCTCA AGGGTGCATG CTCTTGTTGGT TCTTGCTGCA
  25140      25150      25160      25170      25180      25190      25200

          >< FokI
          >< DdeI
>< MnlI >< P1eI>< HinfI >< BsrI
AGTTTGATGA GGATGACTCT GAGCCAGTTC TCAAGGGTGT CAAATTACAT TACACATAAA CGAACTTATG
  25210      25220      25230      25240      25250      25260      25270

  >< Sau3AI
  >< NdeII
  >< MboI
  >< DpnII
    > < DpnI

```

FIGURE 13.58

72/116

```

      >< BspAI
      > < Bsp143I
      >< BsgI      >< AlwI      >< BsrI      BspWI >
GATTTGTTTA TGAGATTTTT TACTCTTGGA TCAATTACTG CACAGCCAGT AAAAATTGAC AATGCTTCTC
25280      25290      25300      25310      25320      25330      25340

      >< ScaI
      >< RsaI
      >< Csp6I      >< SfcI
      >< AfaI      >< NlaIII      >< AciI      >< MnlI      FokI >
CTGCAAGTAC TGTTCATGCT ACAGCAACGA TACCGCTACA AGCCTCACTC CCTTTCGGAT GGCTTGTTAT
25350      25360      25370      25380      25390      25400      25410

      > < HinPII
      > < Hin6I
      >< HhaI
      >< HaeII      >< HinPII      RmaI ><
      >< Eco47III      >< Hin6I      NheI ><
      >< CfoI      >< HhaI      MaeI ><
      >< BspWI      >< Bsp143II      >< CfoI      Fnu4HI ><
      >< BspWI      >< Bsp143II      >< CfoI      AluI ><
TGGCGTTGCA TTTCTTGCTG TTTTTCAGAG CGCTACCAAA ATAATTGCGC TCAATAAAAAG ATGGCAGCTA
25420      25430      25440      25450      25460      25470      25480

      >< EcoNI
      >< BslI
      >< BsiYI
      >< BbvI      >< BsrI      >< BbvI      > < Fnu4HI      BbvI ><
      >< BbvI      >< BsrI      >< BbvI      > < Fnu4HI      BbvI ><
GCCCTTTATA AGGGCTTCCA GTTCATTTGC AATTTACTGC TGCTATTTGT TACCATCTAT TCACATCTTT
25490      25500      25510      25520      25530      25540      25550

      >< BspMI      >< MnlI      >< CfoI      >< AfaI      >< MnlI      >< SfaNI
      >< BspMI      >< MnlI      >< CfoI      >< AfaI      >< MnlI      >< SfaNI
TGCTTGTCGC TGCAGGTATG GAGGCGCAAT TTTTGTACCT CTATGCCTTG ATATATTTTC TACAATGCAT
25560      25570      25580      25590      25600      25610      25620

      >< SfaNI
      >< NspI
      >< NspHI
      >< NlaIII
      >< SfaNI
CAACGCATGT AGAATTATTA TGAGATGTTG GCTTTGTTGG AAGTGCAAAT CCAAGAACCC ATTACTTTAT
25630      25640      25650      25660      25670      25680      25690

      >< Bst1107I
      >< AccI      MaeIII ><
GATGCCAACT ACTTTGTTTG CTGGCACACA CATAACTATG ACTACTGTAT ACCATATAAC AGTGTCACAG
25700      25710      25720      25730      25740      25750      25760

      >< MboII
      >< BstXI ><
      >< MunI >< MaeIII >< MaeIII >< Eco57I >< BbsI MnlI >
      >< MunI >< MaeIII >< MaeIII >< Eco57I >< BbsI MnlI >
ATACAATTGT CGTTACTGAA GGTGACGGCA TTTCAACACC AAAACTCAAA GAAGACTACC AAATTGGTGG
25770      25780      25790      25800      25810      25820      25830

      >< RsaI
      > < NlaIII
      >< HphI
      >< Tru9I >< Tth111I >< Csp6I
      >< DdeI      >< DdeI      >< MseI >< AspI      >< AfaI

```

FIGURE 13.59

```

TTATTCTGAG GATAGGCACT CAGGTGTTAA AGACTATGTC GTTGTACATG GCTATTTTCAC CGAAGTTTAC
25840      25850      25860      25870      25880      25890      25900

                > < HinfI>< P1eI                >< BsrI                Tru9I ><
                >< AluI >< AccI                >< SfcI >< AlwNI                MseI ><
TACCAGCTTG AGTCTACACA AATTACTACA GACACTGGTA TTGAAAATGC TACATTCTTC ATCTTTAACA
25910      25920      25930      25940      25950      25960      25970

                > < TthHB8I
                >< Tru9I                >< TaqI                >< Ksp632I
                >< MseI                >< MboII                >< EarI BspWI ><
>< AluI                >< Eco57I                >< Eam1104I AlwI ><
AGCTTGTTAA AGACCCACCG AATGTGCAAA TACACACAAT CGACGGCTCT TCAGGAGTTG CTAATCCAGC
25980      25990      26000      26010      26020      26030      26040

>< XhoII
>< Sau3AI
>< NlaIV
>< NdeII
>< MflI
>< MboI
>< DpnII
>< DpnI
>< BstYI
>< BstI
>< BspAI
>< Bsp143I
>< BscBI
>< BamHI >< AlwI
AATGGATCCA ATTTATGATG AGCCGACGAC GACTACTAGC GTGCCTTTGT AAGCACAAGA AAGTGAGTAC
26050      26060      26070      26080      26090      26100      26110

                > < Tru9I
                >< RsaI
                > < MseI
                >< MboII
                >< MaeII                >< RsaI
                >< RsaI                >< Csp6I ><
                >< Csp6I                >< Csp6I                >< Tru9I >< Csp6I
                > < AfaI                >< AfaI                >< MseI                >< AfaI
GAACTTATGT ACTCATTCGT TTCGGAAGAA ACAGGTACGT TAATAGTTAA TAGCGTACTT CTTTTTCTTG
26120      26130      26140      26150      26160      26170      26180

                >< TthHB8I
                >< TaqI
                >< RmaI                >< HinfI                > < RsaI
                > < MaeIII                >< HinfI                Fnu4HI ><
                >< MaeI >< RmaI                >< HhaI                >< Csp6I
                >< FokI >< MaeI                >< CfoI >< BbvI > < AfaI
CTTTCGTGGT ATTCTTGCTA GTCACACTAG CCATCCTTAC TGCCTTCCA TTGTGTGCGT ACTGCTGCAA
26190      26200      26210      26220      26230      26240      26250

                >< Tru9I
                >< ThaI
                >< MseI                >< MvnI
                >< SspI >< MaeII                >< MseI
                >< HpaI                >< BstUI                Ksp632I >
                >< HindII                >< MaeII >< Bsp50I                >< MboII EarI >
                >< HincII                >< AccI >< AccII                Eam1104I >
TATTGTAAAC GTGAGTTTAA TAAAACCAAC GGTTCACGTC TACTCGCGTG TAAAAAATCT GAACTCTTCT
26260      26270      26280      26290      26300      26310      26320

```

FIGURE 13.60

74/116

```

>< Sau3AI
>< NdeII
>< MboI
>< DpnII
>< MboII>< DpnI
>< XmnI >< BspAI> < Eco57I
>< Asp700I>< Bsp143I
GAAGGAGTTC CTGATCTTCT GGTCTAAACG AACTAACTAT TATTATTATT CTGTTTGGAA CTTTAACATT
26330      26340      26350      26360      26370      26380      26390

>< ScrFI
>< MvaI
>< EcoRII
>< Ecl136I
>< DsaV NlaIV ><
>< BstOI
>< RsaI
>< MnlI
>< Tru9I
>< Csp6I
>< MseI
>< BstNI RmaI ><
>< NlaIII >< AfaI
>< AluI
>< BsiLI MaeI ><
>< ApyIBscBI ><
GCTTATCATG GCAGACAACG GTACTATTAC CGTTGAGGAG CTAAACAAC TCCTGGAACA ATGGAACCTA
26400      26410      26420      26430      26440      26450      26460

>< ScrFI
>< RmaI
>< MvaI
>< MaeI
>< EcoRII
>< Ecl136I
>< DsaV
>< BstOI
>< BstNI
>< BsiLI
>< ApyI >< MaeIII
GTAATAGGTT TCCTATTCCT AGCCTGGATT ATGTTACTAC AATTTGCCTA TTCTAATCGG AACAGGTTTT
26470      26480      26490      26500      26510      26520      26530

>< Pali
>< MscI
>< MnlI >< MaeIII
>< HaeIII
>< EaeI
>< BsuRI
>< BsrI
>< RsaI
>< Csp6I >< HindIII
>< AfaI >< AluI
>< BspWI
>< BshI
>< BalI
>< BbvI Fnu4HI ><
TGACATAAT AAAGCTTGTT TTCCTCTGGC TCTTGTGGCC AGTAACACTT GCTTGTTTTG TGCTTGCTGC
26540      26550      26560      26570      26580      26590      26600

>< VspI
>< Tru9I
>< MseI
>< SfcI >< AsnI
>< AccI >< AseI>< MaeIII>< AciI
>< HphI
>< BsrI
>< BspI
>< Bsp6I
>< Csp6I
>< Eco57I
>< DdeI
>< CelII
>< Bpu1102I
>< RsaI
>< Csp6I

```

FIGURE 13.61

```

>< BfrI                                     >< AfaI
  >< AluI                                     >< AciI                                     MboII >
CTTAGCTACT TCGTTGCTTC CTTCAGGCTG TTTGCTCGTA CCCGCTCAAT GTGGTCATTC AACCCAGAAA
  26680      26690      26700      26710      26720      26730      26740

      >< ScrFI
      >< NciI
      >< MspI
      >< HpaII
      >< HapII
      >< DsaV>< MnlI
      >< BslI
      >< BsiYI
      >< BsaJI >< MunI          > < XcmI
      >< BcnI          >< MaeIII >< AciI >< NlaIII
CAAACATTCT TCTCAATGTG CCTCTCCGGG GGACAATTGT GACCAGACCG CTCATGGAAA GTGAACCTGT
  26750      26760      26770      26780      26790      26800      26810

Tru9I ><
      SinI >
      Sau96I >
      PpuMI >
      NspIV >
      MseI ><
      >< MaeIII
      >< Sau3AI          > < RmaI >< HaeII
      >< NdeII          >< Pali          > < MaeI          Eco0109I >
      >< MboI          >< MspI          >< HinPIIEco47I >
      >< FbaI          >< HpaII          >< StyI>< Hin6I   DraII >
      >< DpnII          >< HapII          >< EcoT14I          Cfr13I >
      >< DpnI          >< HaeIII          >< Eco130I>< Bsp143II
      >< BspAI          >< GdiII          >< BssT1I          BsiZI >
      >< Bsp143I          >< EaeI          >< BsaJI          Bme18I >
      >< BsiQI          >< BsuRI          >< BlnI          >< HhaI   AvaII >
      >< BclI          >< MaeIII          >< BshI          >< AvrII >< CfoI   AsuI >
CATTGGTGCT GTGATCATTC GTGGTCACTT GCGAATGGCC GGACACTCCC TAGGGCGCTG TGACATTAAG
  26820      26830      26840      26850      26860      26870      26880

      >< Sau3AI
      >< NdeII
      >< MboI
      >< DpnII
      >< DpnI
      >< PssI >< BspMI
      >< Psp5II          >< BspAI          >< XmnI
      >< NspHII          >< Bsp143I          >< Asp700I > < HgaI          Fnu4HI ><
GACCTGCCAA AAGAGATCAC TGTGGCTACA TCACGAACGC TTTCTTATTA CAAATTAGGA GCGTCGCAGC
  26890      26900      26910      26920      26930      26940      26950

      >< TfiI
      >< HinfI
      >< BbvI
      >< BbvI          >< Fnu4HI >< AciI          > < Tru9I
      >< BbvI          >< Fnu4HI >< AciI          > < MseI
GTGTAGGCAC TGATTACAGGT TTTGCTGCAT ACAACCGCTA CCGTATTGGA AACTATAAAT TAAATACAGA
  26960      26970      26980      26990      27000      27010      27020

      >< MspI          >< RsaI
      >< HpaII          >< RmaI
      >< HapII          >< Csp6I
      >< Cfr10I          >< MaeI>< BcgI          HindII ><
      >< BcgI/a          >< SspI          >< AfaI >< MaeIII          HincII ><

```

FIGURE 13.62

76/116

```

CCACGCCGGT AGCAACGACA ATATTGCTTT GCTAGTACAG TAAGTGACAA CAGATGTTTC ATCTTGTTGA
  27030      27040      27050      27060      27070      27080      27090

  >< ScrFI
  >< MvaI
    >< MaeIII
  >< EcoRII
  >< Ecl136I
  >< DsaV
  >< BstOI
  >< BstNI
  >< BsiLI
  >< ApyI
                                >< MnlI
                                HinfI ><
CTTCCAGGTT ACAATAGCAG AGATATTGAT TATCATTATG AGGACTTTCA GGATTGCTAT TTGAATCTT
  27100      27110      27120      27130      27140      27150      27160

                                >< BsmAI
                                >< Tru9I
                                > < MnlI
  >< MaeII
                                >< Alw26I
                                >< MseI
                                >< DdeI
                                >< MboII
GACGTTATAA TAAGTTCAAT AGTGAGACAA TTATTTAAGC CTCTAACTAA GAAGAATTAT TCGGAGTTAG
  27170      27180      27190      27200      27210      27220      27230

                                >< Ksp632I
                                >< MboII
                                >< EarI
                                >< NlaIIIEam1104I ><
ATGATGAAGA ACCATGGAG TTAGATTATC CATAAACGA ACATGAAAAT TATTCTCTTC CTGACATTGA
  27240      27250      27260      27270      27280      27290      27300

                                > < RsaI >< RsaI
                                >< Csp6I >< Csp6I
                                > < AluI
                                >< MnlI
                                > < AfaI >< AfaI
TTGTATTTAC ATCTTGCGAG CTATATCACT ATCAGGAGTG TGTTAGAGGT ACGACTGTAC TACTAAAAGA
  27310      27320      27330      27340      27350      27360      27370

                                >< MnlI >< HphI >< HphI
                                >< MnlI
ACCTTGCCCA TCAGGAACAT ACGAGGGCAA TTCACCATTT CACCCTCTTG CTGACAATAA ATTTGCACTA
  27380      27390      27400      27410      27420      27430      27440

                                Sau3AI >
                                > < PvuII
                                > < Psp5I
                                > < NspBII
                                >< TthHB8I
                                >< TaqI
                                >< RsaI
                                >< Csp6I
                                >< BbvI
                                >< AfaI
                                > < AluI
  >< RmaI
  >< MaeI
ACTTGCACTA GCACACACTT TGCTTTTGCT TGTGCTGACG GTACTCGACA TACCTATCAG CTGCGTGCAA
  27450      27460      27470      27480      27490      27500      27510

                                >< SstI
                                >< SduI
                                >< SacI
                                >< NspII
                                >< HgiAI
                                >< Eco24I
                                > < Ecl136II
                                >< BspWI
                                >< Bsp1286I
                                >< BmyI
                                >< BanII
                                >< Alw21I

  >< HphI
  >< DpnI
                                >< MnlI

```

FIGURE 13. 63

77/116

```

>< Bsp143I          >< MnlI          > < AluI      BbvI ><
GATCAGTTTC ACCAAAACTT TTCATCAGAC AAGAGGAGGT TCAACAAGAG CTCTACTCGC CACTTTTTTCT
  27520      27530      27540      27550      27560      27570      27580

                                     SstI ><
                                     SduI ><
                                     SacI ><
                                     NspII ><
                                     HgiAI ><
                                     Eco24I ><
                                     Ecl136II ><
                                     Bsp1286I ><
                                     BmyI ><
                                     BanII ><
                                     Alw21I ><
                                     AluI ><
>< Fnu4HI          >< HphI          >< Tru9I      >< Tru9I
GCTCTAGTAT TTTTAATACT TTGCTTCACC ATTAAGAGAA AGACAGAATG AATGAGCTCA
  27590      27600      27610      27620      27630      27640      27650

>< Tru9I          >< Tru9I
>< MseI          >< MseI
CTTTAATTGA CTTCTATTTG TGCTTTTTAG CCTTCTGCT ATTCCTTGTT TTAATAATGC TTATTATATT
  27660      27670      27680      27690      27700      27710      27720

>< XhoII
>< XbaI
> < ScrFI
>< Sau3AI
>< RmaI
>< NdeII
> < MvaI
>< MflI
>< MboI
>< EcoRII>< MaeI
> < Ecl136I
>< DpnII
>< DpnI
>< BstYI
> < BstOI
> < BstNI
>< TthHB8I >< BspAI          > < RsaI
>< DsaV>< Bsp143I          >< MboII
> < BsiLI          >< Csp6I
>< TaqI > < ApyI > < AlwI > < AfaI          >< NlaIII
TTGGTTTTCA CTCGAAATCC AGGATCTAGA AGAACCTTGT ACCAAAGTCT AAACGAACAT GAAACTTCTC
  27730      27740      27750      27760      27770      27780      27790

                                     >< HinPII
                                     >< Hin6I
                                     >< HhaI
                                     >< RsaI >< HaeII
                                     >< SfcI          >< Eco47III
                                     >< Csp6I>< CfoI SfaNI ><
                                     >< AfaI >< Bsp143II
>< NdeI
ATTGTTTTGA CTTGTATTTT TCTATGCAGT TGCATATGCA CTGTAGTACA GCGCTGTGCA TCTAATAAAC
  27800      27810      27820      27830      27840      27850      27860

>< XhoII
>< Sau3AI
>< NdeII
> < MnlI
>< MflI

```

FIGURE 13.64

78/116

```

    >< MboI
    >< DpnII
      >< DpnI      >< RsaI
      >< BstYI    >< MboII
    >< NlaIII>< BspAI      >< Csp6I >< RmaI
      >< AlwI >< Bsp143I   >< AfaI >< MaeI
CTCATGTGCT TGAAGATCCT TGTAAGGTAC AACACTAGGG GTAATACTTA TAGCACTGCT TGGCTTTGTG
  27870      27880      27890      27900      27910      27920      27930

>< SduI
  >< RmaI
>< NspII
  >< MaeI
>< HgiAI
>< Bsp1286I
>< BmyI
>< Alw21I
                                     >< NspI
                                     >< NspHI
                                     >< NlaIII >< MaeIII
CTCTAGGAAA GGTTTTACCT TTTCATAGAT GGCACACTAT GGTTCAAACA TGCACACCTA ATGTTACTAT
  27940      27950      27960      27970      27980      27990      28000

  > < XhoII
  > < Sau3AI > < Van91I
                                     >< RsaI
                                     >< NlaIV
                                     >< KpnI >< NlaIII
  > < NdeII > < PflMI
                                     >< Eco64I >< MaeIII
  > < MflI>< NspBII
                                     >< Csp6I>< HphI
  > < DpnII      >< HinPII
                                     >< BscBI >< EcoO65I
      >< Bsp143I   >< Hin6I
                                     >< BanI >< BspHI
  > < BstYI > < BslI >< HhaI >< RmaI >< Asp718 >< Eco91I
  > < BspAI > < BsiYI>< CfoI >< MaeI >< AfaI >< BstPI
  > < MboI>< AluI>< BspWI >< BspWI >< AccB1I >< BstEII
    >< AlwI >< DpnI > < AccB7I >< AluI >< Acc65I >< BbvI
CAACTGTCAA GATCCAGCTG GTGGTGCGCT TATAGCTAGG TGTTGGTACC TTCATGAAGG TCACCAAAC
  28010      28020      28030      28040      28050      28060      28070

                                     >< SinI
                                     >< Sau96I
                                     >< NspIV
NspHII ><
  NlaIV ><
                                     >< Eco47I
                                     >< Cfr13I
                                     >< BsiZI
BscBI ><
  >< Esp3I >< Csp6I >< Tru9I >< Bme18I
  >< BsmAI >< BsmBI >< MseI >< Tru9I >< AvaII
  >< Alw26I >< AfaI >< DraI >< MseI >< AsuI
GCTGCATTTA GAGACGTACT TGTTGTTTAA AATAAACGAA CAAATTAAAA TGTCTGATAA TGGACCCCAA
  28080      28090      28100      28110      28120      28130      28140

                                     >< SinI
                                     >< Sau96I
                                     >< NspIV
                                     >< NspHII
                                     >< NlaIV
                                     >< Eco47I
                                     >< Cfr13I
                                     >< BsiZI
      >< SduI
      >< NspII
      >< Bsp1286I
      >< BmyI
    >< MaeII >< AciI
                                     >< BscBI
                                     >< Bme18I
                                     >< AvaII >< TfiI
                                     >< AsuI >< HinfI
                                     >< MnlI

```

FIGURE 13. 65

79/116

```

TCAAACCAAC GTAGTGCCCC CCGCATTACA TTTGGTGGAC CCACAGATTC AACTGACAAT AACCAGAATG
28150      28160      28170      28180      28190      28200      28210

      >< HinP1I >< StyI
      >< HaeII
      > < Pali >< Hin6I >< EcoT14I
      > < HaeIII >< HhaI>< Eco130I
      >< BspWI      >< BssT1I
      > < BsuRI      >< Bsp143II
      >< HgaI> < BshI      >< CfoI>< BsaJI      >< HgaI
GAGGACGCAA TGGGGCAAGG CCAAAACAGC GCCGACCCCA AGGTTTACCC AATAATACTG CGTCTTGTT
28220      28230      28240      28250      28260      28270      28280

      >< TthHB8I
      > < ScrFI
      >< Pali
      >< PaeR7I
      >< NspIII
      > < MvaI
      >< HaeIII
      >< EcoRII
      >< Eco88I
      >< XhoI > < Ecl136I
      >< DsaV
      >< BsuRI
      >< SlaI > < BstOI
      >< MnlI>< TaqI> < BstNI
      >< CcrI > < BsiLI
      >< HinfI      >< BshI
      >< TfiI>< BcoI>< BsaJI
      >< MnlI      >< DdeI      >< AvaI > < ApyI
      >< AluI >< DdeI > < NlaIII      >< BfrI      >< Ama87I >< MnlI
CACAGCTCTC ACTCAGCATG GCAAGGAGGA ACTTAGATTC CCTCGAGGCC AGGGCGTTCC AATCAACACC
28290      28300      28310      28320      28330      28340      28350

      >< SinI
      >< Sau96I
      >< NspIV
      >< NspHII
      >< Eco47I
      >< Cfr13I
      >< BsiZI
      >< Bme18I      > < Ksp632I
      >< AvaII      > < Eam1104I
      >< AsuI      > < EarI      > < AluI>< MboII      >< MaeIII
AATAGTGGTC CAGATGACCA AATTGGCTAC TACCGAAGAG CTACCCGACG AGTTCGTGGT GGTGACGGCA
28360      28370      28380      28390      28400      28410      28420

      >< SstI
      >< SduI
      >< SacI
      >< NspII
      >< HgiAI
      >< EspI
      >< Eco24I
      >< Ecl136II      >< StyI      >< Sau96I
      >< DdeI      >< RmaI      >< Pali
      >< CelII      >< MaeI      >< NspIV
      >< Bsp1286I      >< EcoT14I >< HaeIII
      >< Bpu1102I      >< Eco130I >< Cfr13I
      >< BmyI      >< BssT1I >< BsuRI
      >< BanII      >< BsrI
      >< RsaI      >< BsaJI      >< BsiZI

```

FIGURE 13. 66

80/116

```

      >< Alw21I      >< Csp6I      >< BlnI      >< BshI>< HindIII
>< HphI  >< AluI      >< AfaI      >< AvrII      >< AsuI      >< AluI
AAATGAAAGA GCTCAGCCCC AGATGGTACT TCTATTACCT AGGAAGCTGGC CCAGAAGCTT CACTTCCCTA
28430      28440      28450      28460      28470      28480      28490

>< HinPII
>< Hin6I
  >< HhaI
    >< HaeII
  >< CfoI      > < MnlI      >< NlaIV
    >< Bsp143II      >< SfaNI >< DdeI >< BscBI
CGGCGCTAAC AAAGAAGGCA TCGTATGGGT TGCAACTGAG GGAGCCTTGA ATACACCCAA AGACCACATT
28500      28510      28520      28530      28540      28550      28560

>< NlaIV
>< Eco64I
  >< BscBI
>< BanI
  >< AciI
>< AccB1I >< BbvI      >< Fnu4HI      >< MnlI
GGCACCCGCA ATCCTAATAA CAATGCTGCC ACCGTGCTAC AACTTCCTCA AGGAACAACA TTGCCAAAAG
28570      28580      28590      28600      28610      28620      28630

      >< ThaI
      >< MnlI
      >< MaeII >< MvnI
      >< MnlI      BstUI ><
      >< Fnu4HI      >< Ksp632I      Bsp50I ><
      >< BspWI      >< EarI      >< BsaAI>< AciI
>< MnlI      >< MnlI      >< AciI>< MboII      >< Eam1104I      AccII ><
GCTTCTACGC AGAGGGAAGC AGAGGCGGCA GTCAAGCCTC TTCTCGCTCC TCATCACGTA GTCGCGGTAA
28640      28650      28660      28670      28680      28690      28700

      >< ScrFI
      >< MvaI
      >< EcoRII
      >< Ecl136I
      >< DsaV>< Fnu4HI
      >< BstOI
      >< BstNI
      >< BsiLI
      >< ApyI      >< BbvI      >< TaqI      >< AciI
TTCAAGAAAT TCAACTCCTG GCAGCAGTAG GGGAAATTCT CCTGCTCGAA TGGCTAGCGG AGGTGGTGAA
28710      28720      28730      28740      28750      28760      28770

      > < ThaI
      > < MvnI
>< HphI >< MnlI
  > < HinPII
  > < Hin6I
    >< HhaI
      > < BstUI      >< RmaI
      > < Bsp50I      >< MaeI
>< BbvI >< CfoI>< Fnu4HI
  > < AccII>< BspWI      >< AluI
ACTGCCCTCG CGCTATTGCT GCTAGACAGA TTGAACCAGC TTGAGAGCAA AGTTTCTGGT AAAGGCCAAC
28780      28790      28800      28810      28820      28830      28840

      >< PalI>< MaeIII
      >< HaeIII
      >< BsuRI      >< DdeI
      >< Fnu4HI
      >< DdeI
      >< RsaI ><
      >< MnlI
      >< MaeII ><
      >< Csp6I ><

```

FIGURE 13.67

81/116

```

      > < BshI   > < BbvI   >< MnlI >< BspWI   >< SfaNI   AfaI ><
AACAAACAAGG CCAAACTGTC ACTAAGAAAT CTGCTGCTGA GGCATCTAAA AAGCCTCGCC AAAAACGTAC
      28850      28860      28870      28880      28890      28900      28910

                                >< Tth111I
                                >< SinI
                                >< Sau96I
                                >< NspIV
                                >< NspHII
      > < MaeII
                                >< Eco47I
                                >< Cfr13I
                                >< BsmBI
                                >< BsiZI   >< StyI
                                >< Bme18I   >< EcoT14I
                                >< AvaII    >< Eco130I
                                >< AsuI     >< BssT1I
                                >< BsaJI
      TGCCACAAAA CAGTACAACG TCACTCAAGC ATTTGGGAGA CGTGGTCCAG AACAAACCCA AGGAAATTTTC
      28920      28930      28940      28950      28960      28970      28980

      >< SinI
      >< Sau96I
      >< NspIV
      >< NspHII
      >< NlaIV
      >< Eco47I
      >< Cfr13I
      >< BsiZI
      >< BscBI
      >< Bme18I
      >< AvaII
      >< AsuI
      GGGGACCAAG ACCTAATCAG ACAAGGAACT GATTACAAAC ATTGGCCGCA AATTGCACAA TTTGCTCCAA
      28990      29000      29010      29020      29030      29040      29050

                                >< Pali
                                >< HaeIII
      >< GdiII
                                >< Fnu4HI
      >< EaeI
                                >< BsuRI
                                >< BshI
                                >< AciI
                                BspWI >
                                >< BspWI
      GGGGACCAAG ACCTAATCAG ACAAGGAACT GATTACAAAC ATTGGCCGCA AATTGCACAA TTTGCTCCAA
      28990      29000      29010      29020      29030      29040      29050

      >< BsmI
      >< BscCI >< MnlI >< MaeIII
                                >< NlaIII
                                >< MaeIII
                                >< NlaIII
      GTGCCTCTGC ATTCTTTGGA ATGTCACGCA TTGGCATGGA AGTCACACCT TCGGGAACAT GGCTGACTTA
      29060      29070      29080      29090      29100      29110      29120

                                >< XhoII
                                >< Sau3AI
                                >< NdeII
                                >< MflI
                                >< MboI
                                >< FokI
      >< Tru9I
      >< NlaIV
      >< NlaIII
      >< MseI
      >< BscBI >< BstXI>< AlwI> < Bsp143I
                                >< DpnII
                                > < DpnI
                                >< BstYI
                                >< BspAI
                                >< Tth111I
                                >< MaeII
                                BspWI ><
      TCATGGAGCC ATTAAATTGG ATGACAAAGA TCCACAATTC AAAGACAACG TCATACTGCT GAACAAGCAC
      29130      29140      29150      29160      29170      29180      29190

                                EspI ><
                                DdeI ><
                                CelII ><
                                Bpu1102I ><
                                AluI ><
      >< HgaI
      ATTGACGCAT ACAAACATT CCCACCAACA GAGCCTAAAA AGGACAAAAA GAAAAAGACT GATGAAGCTC
      29200      29210      29220      29230      29240      29250      29260

```

FIGURE 13.68


```

                >< PleI
    >< Fnu4HI
    >< BspWI
    >< BsmAI
    >< Alw26I
    >< AciI
AGCCTTTGCC GCAGAGACAA AAGAAGCAGC CCACTGTGAC TCTTCTTCCT GCGGCTGACA TGGATGATTT
    29270      29280      29290      29300      29310      29320      29330

                >< Fnu4HI >< BbvI >< AciI >< NlaIII
                >< NlaIII >< HinfI >< HinfI >< NlaIII ><
    >< FokI
CTCCAGACAA CTTCAAAATT CCATGAGTGG AGCTTCTGCT GATTCAACTC AGGCATAAAC ACTCATGATG
    29340      29350      29360      29370      29380      29390      29400

                >< MaeII
ACCACACAAG GCAGATGGGC TATGTAAACG TTTTCGCAAT TCCGTTTACG ATACATAGTC TACTCTTGTG
    29410      29420      29430      29440      29450      29460      29470

                >< Tru9I
                >< Tru9I
                >< MseI
    >< MseI
    >< XmnI
    >< EcoRI>< MaeIII
    >< Asp700I >< BsgI
CAGAATGAAT TCTCGTAACT AAACAGCACA AGTAGGTTTA GTTAACTTTA ATCTCACATA GCAATCTTTA
    29480      29490      29500      29510      29520      29530      29540

                >< XorII >
                >< TthHB8I >
                >< TaqI >
                >< Sau3AI ><
                >< RsaI ><
                >< ThaIPvuI >
                >< NdeII ><
                >< MnlI
                >< MvnIMcrI >
                >< MboI ><
                >< DpnII ><
                >< DpnI ><
                >< Csp6I ><
                >< BstUI
                >< HaeIII BspCI >
                >< BspAI ><
                >< TthHB8I >< Bsp50I
                >< Pali Bspl43I ><
                >< BsuRI BsiEI >
                >< BshIAfaI ><
                >< MnlI
                >< MaeIII
ATCAATGTGT AACATTAGGG AGGACTTGAA AGAGCCACCA CATTTTCATC GAGGCCACGC GGAGTACGAT
    29550      29560      29570      29580      29590      29600      29610

                >< SduI
                >< NspII
                >< MboII >< VspI
                >< Ksp632I >< Eco24I >< Tru9I
                >< RsaI >< RmaI >< Fnu4HI >< Bsp1286I >< MseI
                >< Csp6I >< MaeI >< EarI >< BmyI >< AsnI
                >< AfaI >< BbvI >< AluI>< Eam1104I >< BanII >< AseI

```

FIGURE 13.69

CGAGGGTACA GTGAATAATG CTAGGGAGAG CTGCCTATAT GGAAGAGCCC TAATGTGTAA AATTAATTTT
29620 29630 29640 29650 29660 29670 29680

>< Tru9I >< DdeI

>< MseI >< BfrI

>< NlaIII > < AluI

AGTAGTGCTA TCCCCATGTG ATTTTAATAG CTTCTTAGGA GAATGACAAA AAAAAAAAAA AAAAAA
29690 29700 29710 29720 29730 29740

Sérologie SRAS : Technique N Indirect (1ère série)

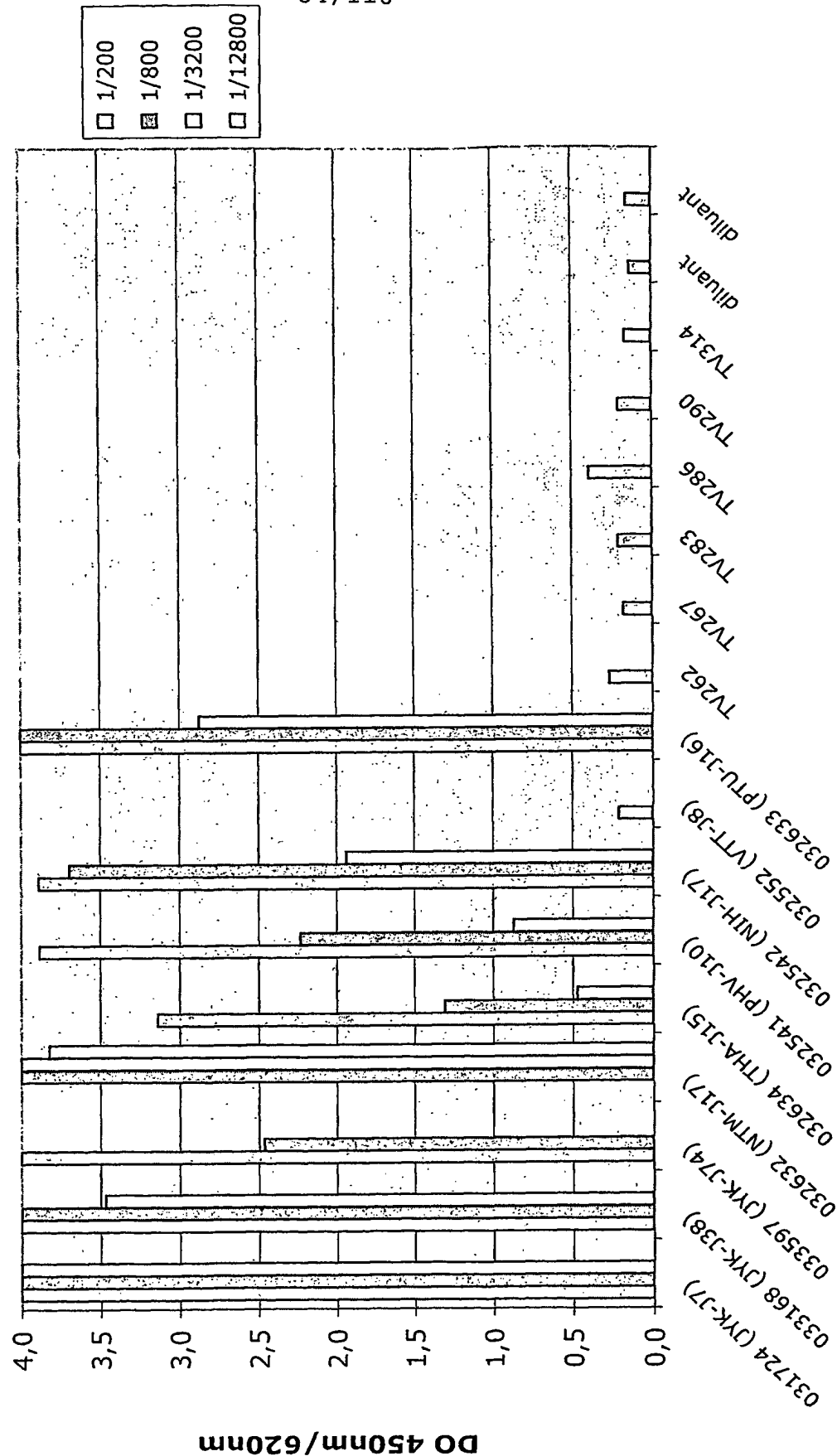


FIGURE 14

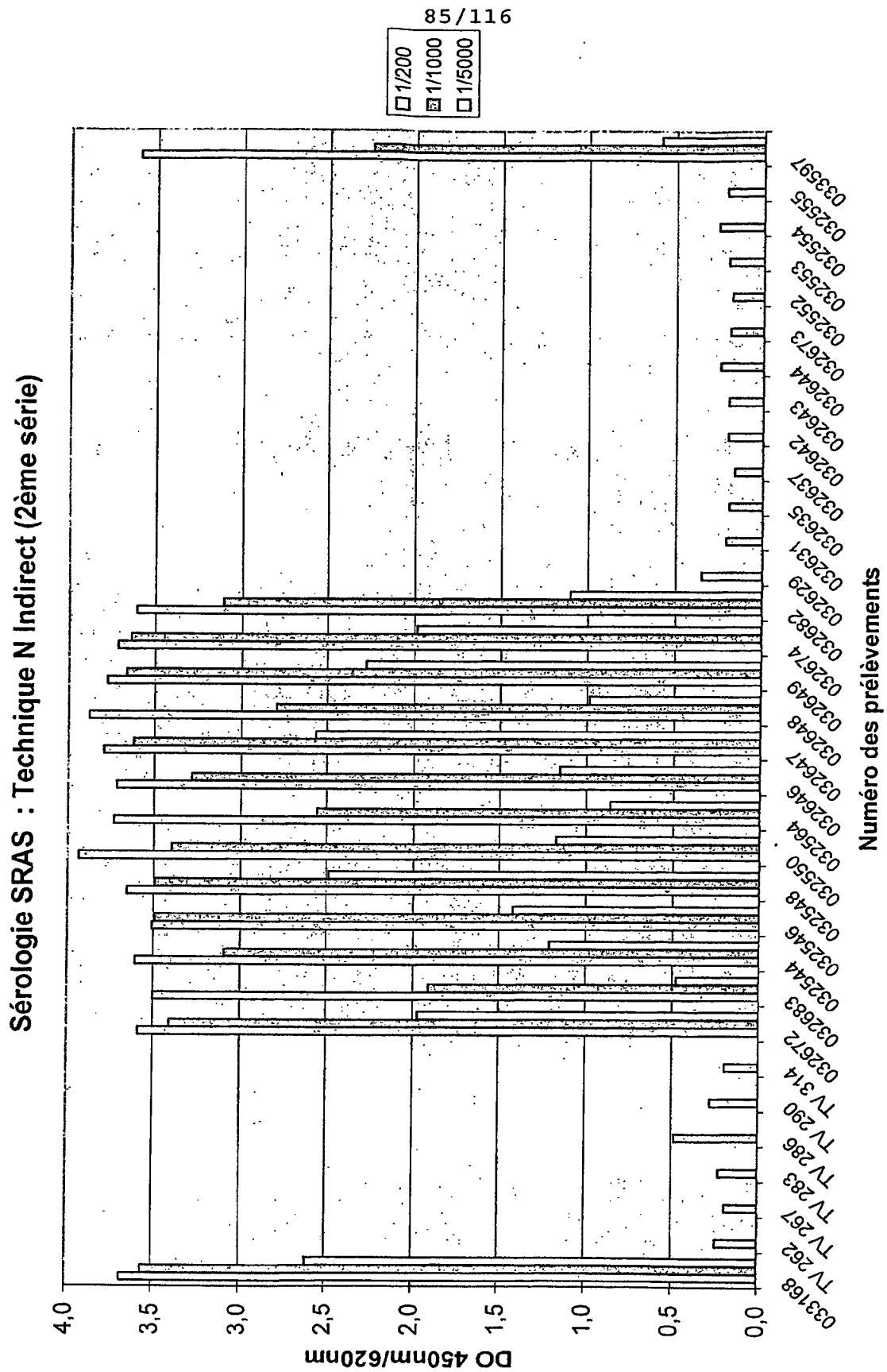


FIGURE 15

Sérologie SRAS : Technique Double Epitope (1ère série)

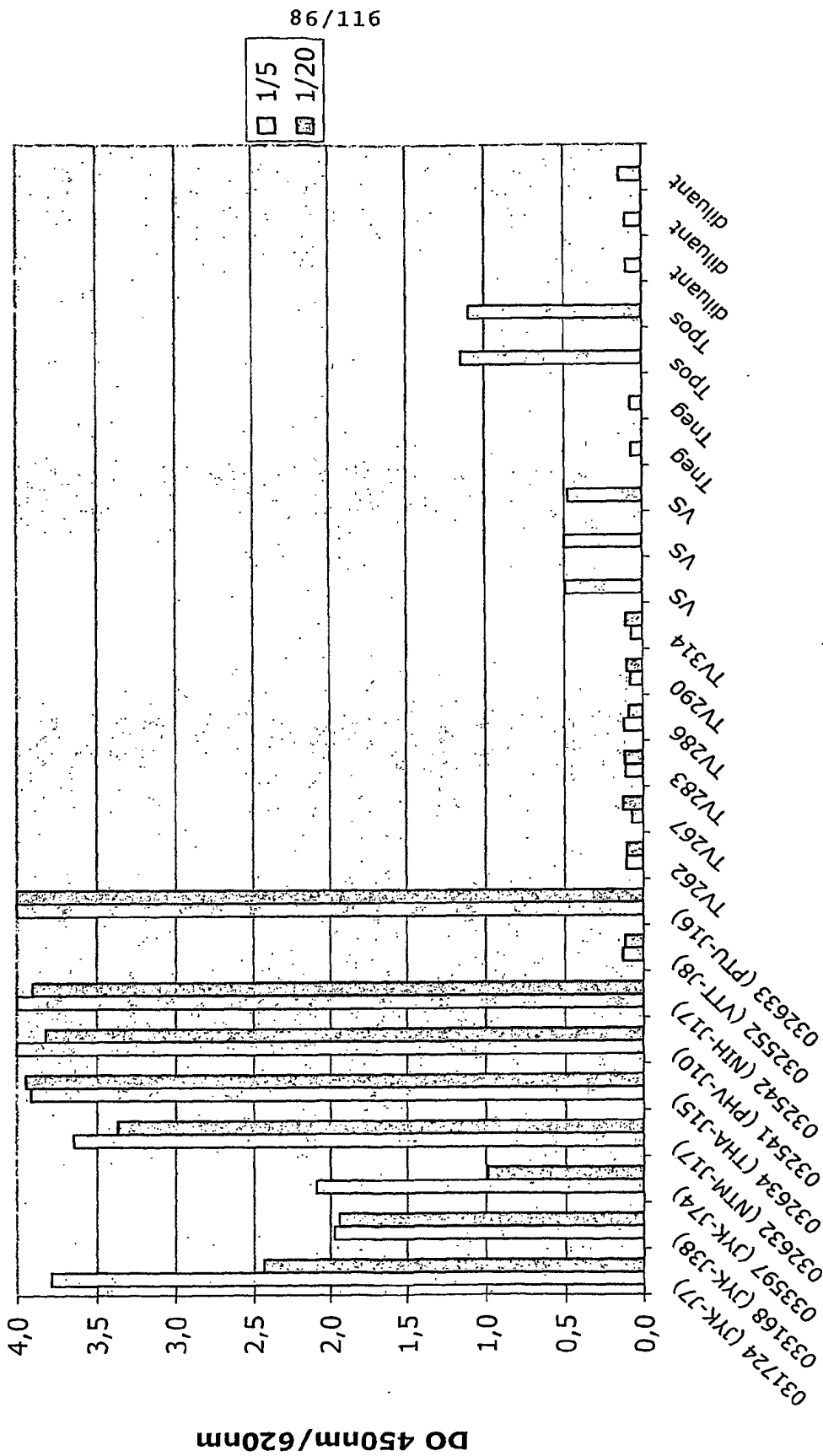


FIGURE 16

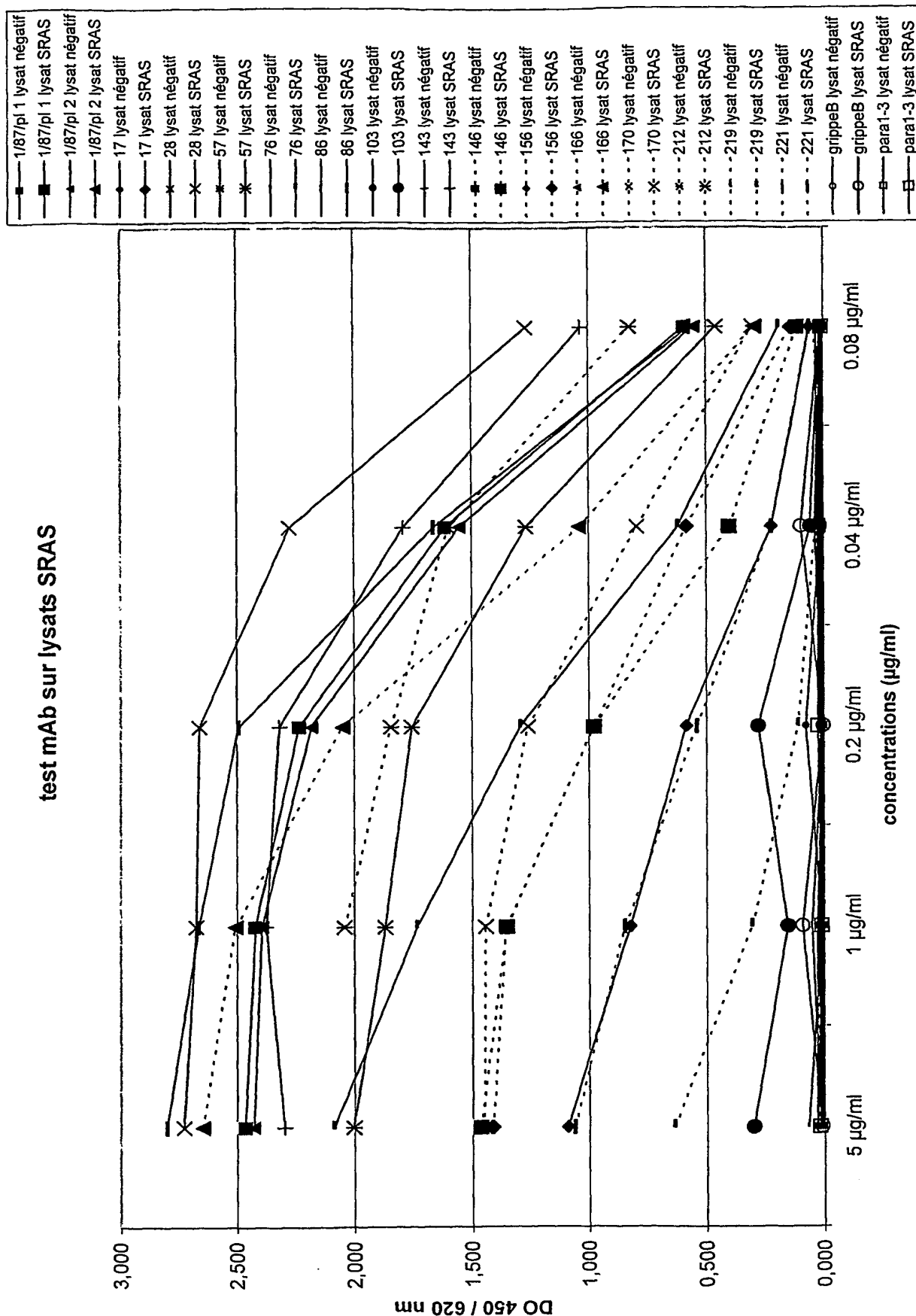


FIGURE 18

Test mAb sur lysats HCoV-229E

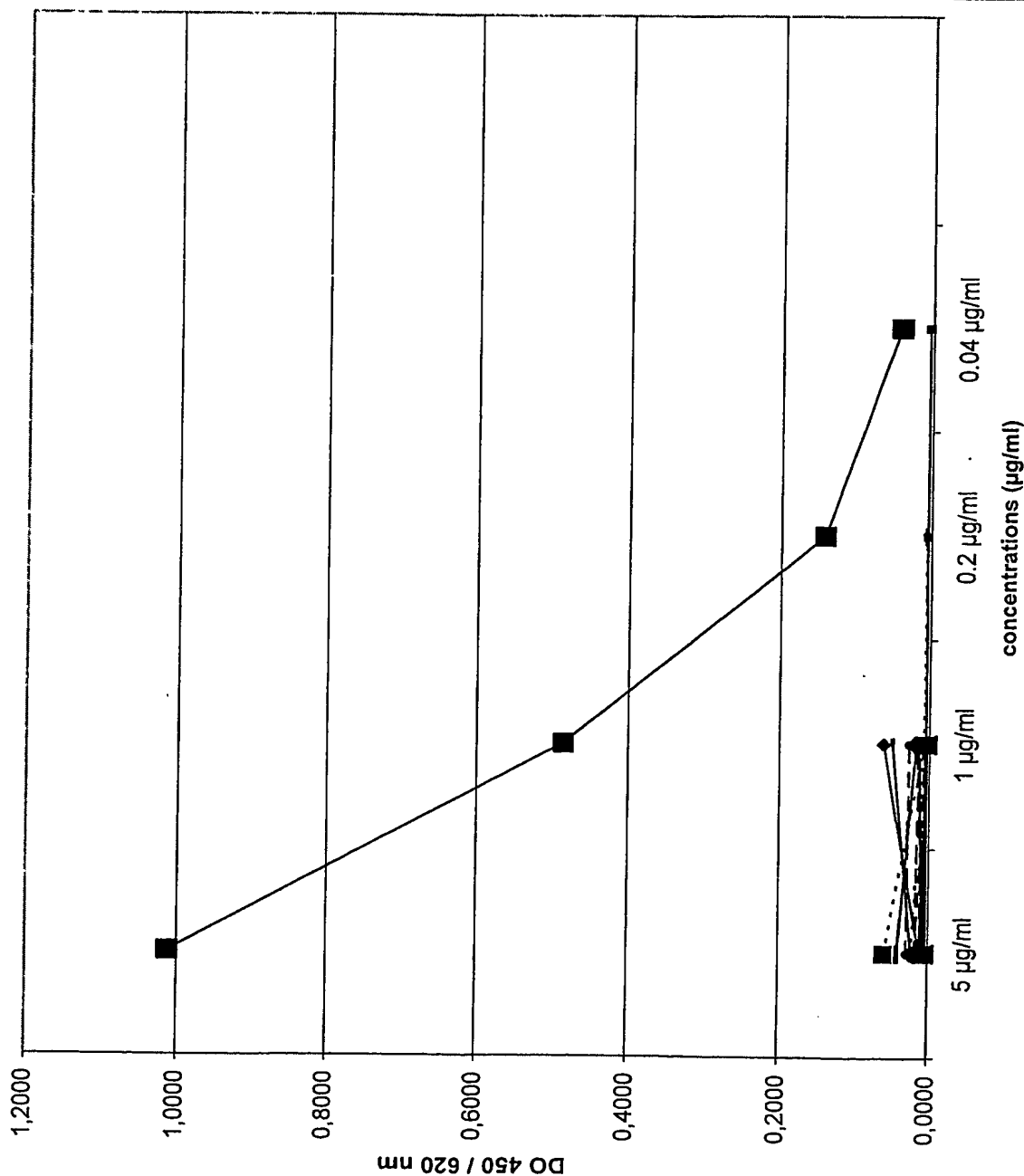


FIGURE 19

- ▲— 1/87 lysat négatif
- ▲— 1/87 lysat 229E
- 5-11H.6 lysat négatif
- 5-11H.6 lysat 229E
- 17 lysat négatif
- 17 lysat 229E
- ◆— 28 lysat négatif
- ◆— 28 lysat 229E
- ♦— 57 lysat négatif
- ♦— 57 lysat 229E
- ▲— 76 lysat négatif
- ▲— 76 lysat 229E
- 86 lysat négatif
- 86 lysat 229E
- 103 lysat négatif
- 103 lysat 229E
- ◆— 143 lysat négatif
- ◆— 143 lysat 229E
- ♦— 146 lysat négatif
- ♦— 146 lysat 229E
- ▲— 156 lysat négatif
- ▲— 156 lysat 229E
- 166 lysat négatif
- 166 lysat 229E
- 170 lysat négatif
- 170 lysat 229E
- ◆— 212 lysat négatif
- ◆— 212 lysat 229E
- ♦— 219 lysat négatif
- ♦— 219 lysat 229E
- ▲— 221 lysat négatif
- ▲— 221 lysat 229E
- grippeB lysat négatif
- grippeB lysat 229E
- para1-3 lysat négatif
- para1-3 lysat 229E

90/116

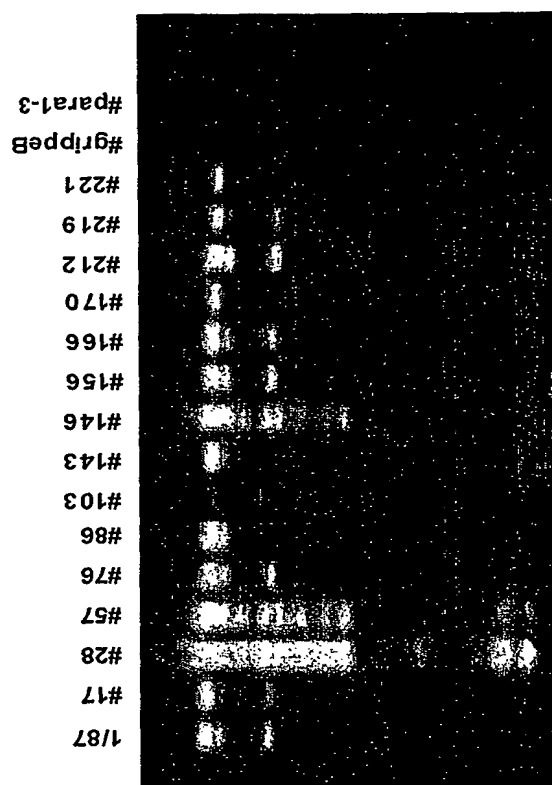


FIGURE 20

91/116

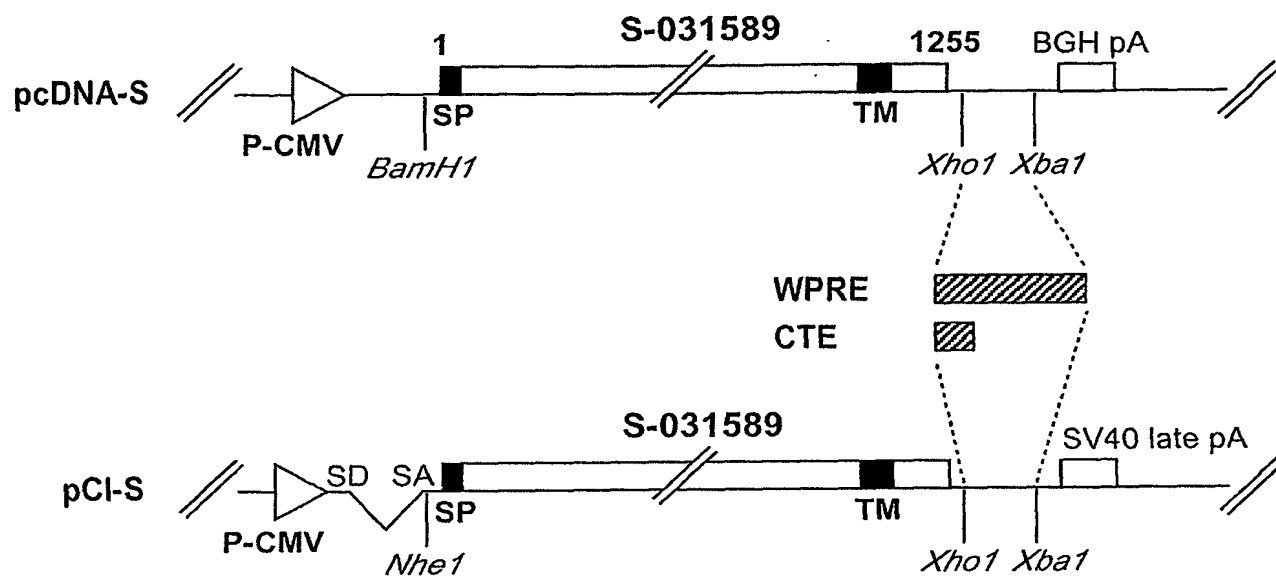


FIGURE 21

92/116

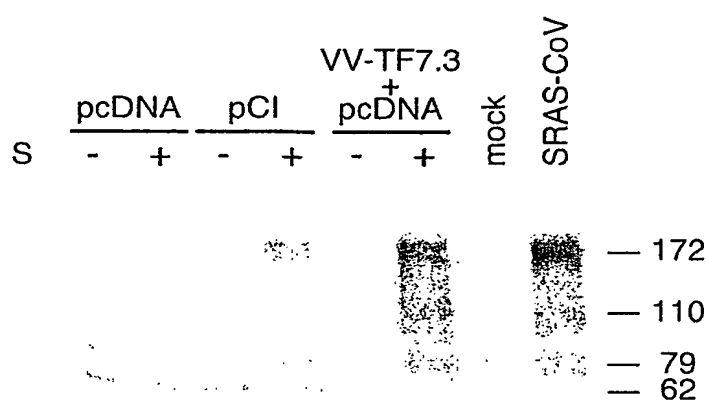
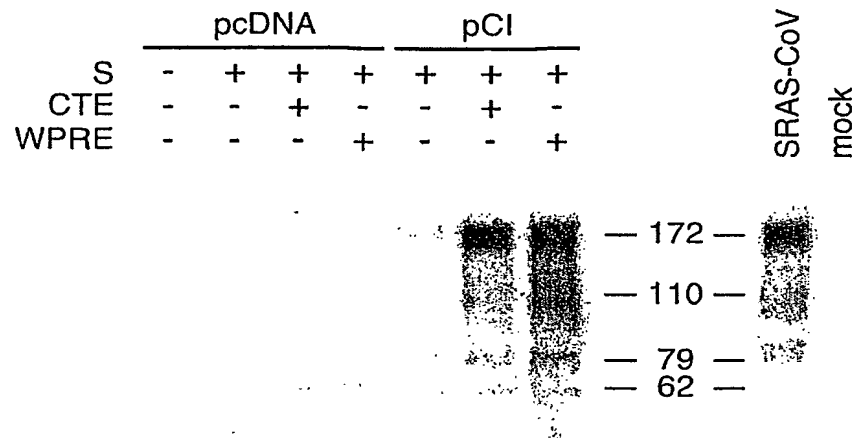


FIGURE 22

93/116

A.



B.

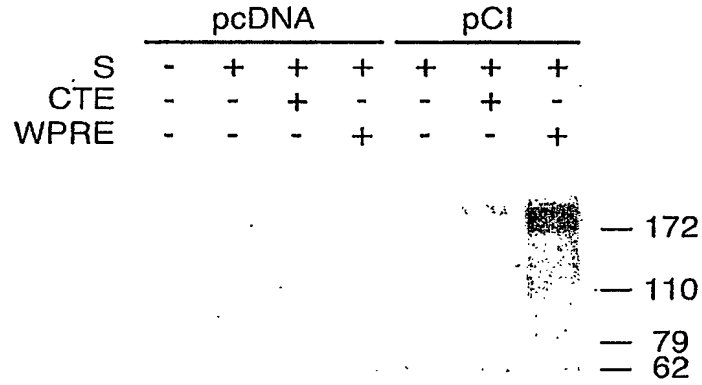


FIGURE 23

94/116

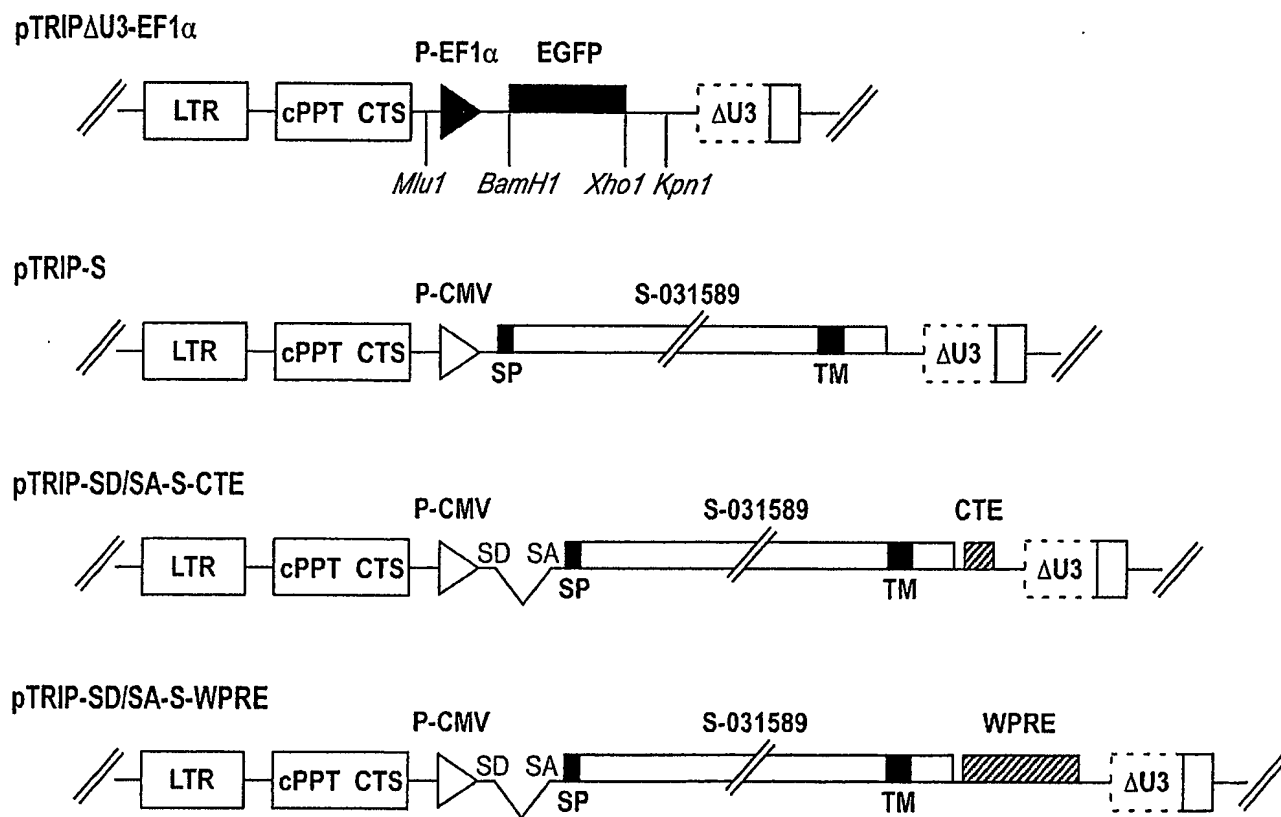


FIGURE 24

95/116

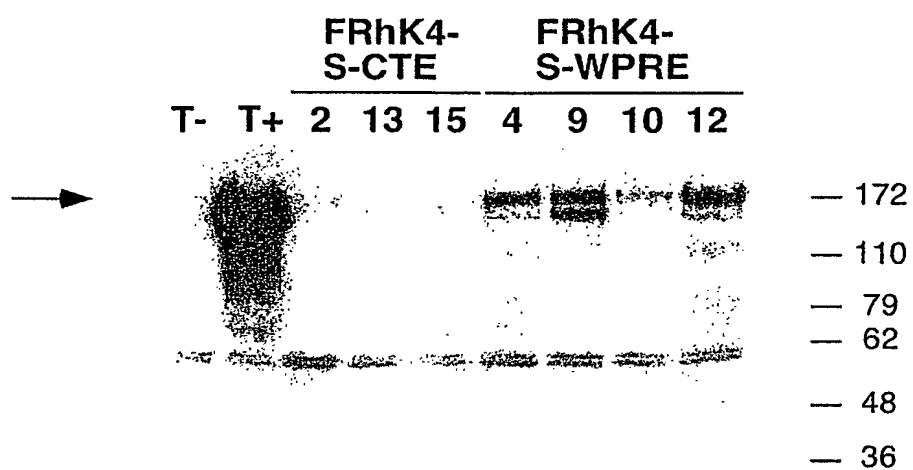


FIGURE 25

96/115

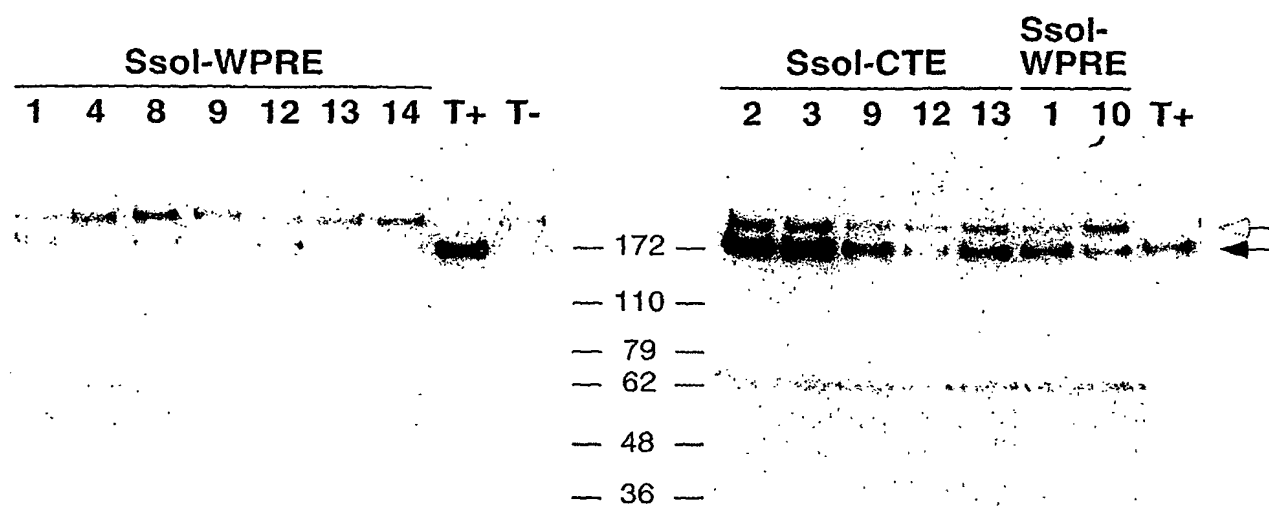
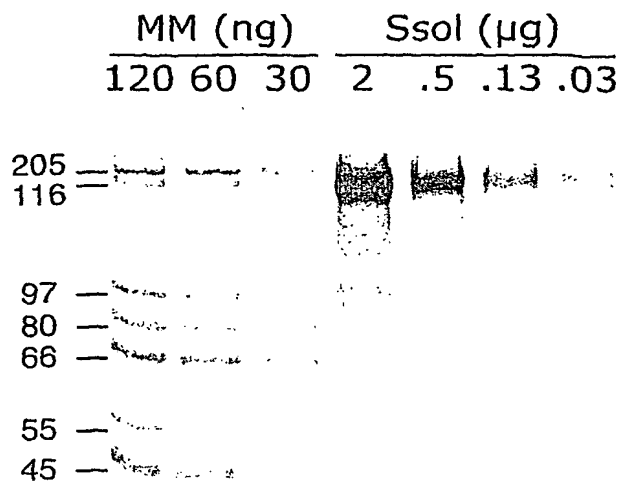


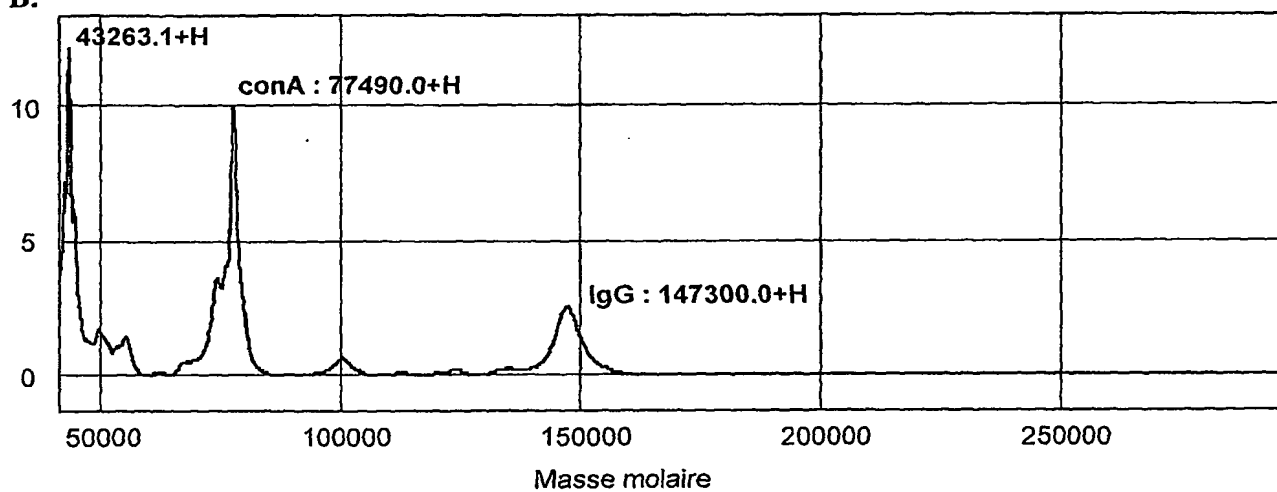
FIGURE 26

97/116

A.



B.



C.

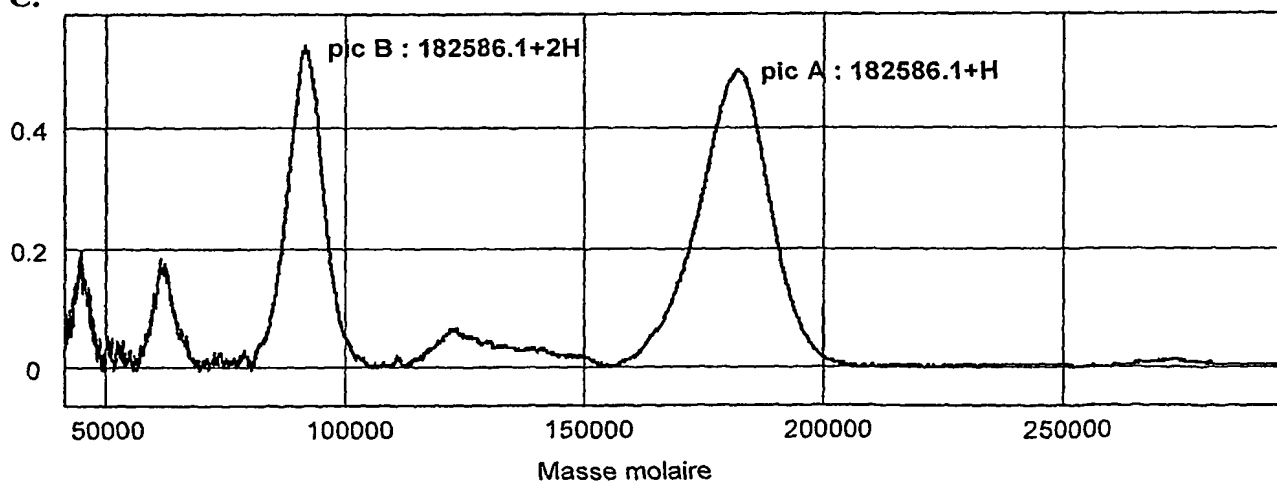


FIGURE 27 A-C

98/116

D.

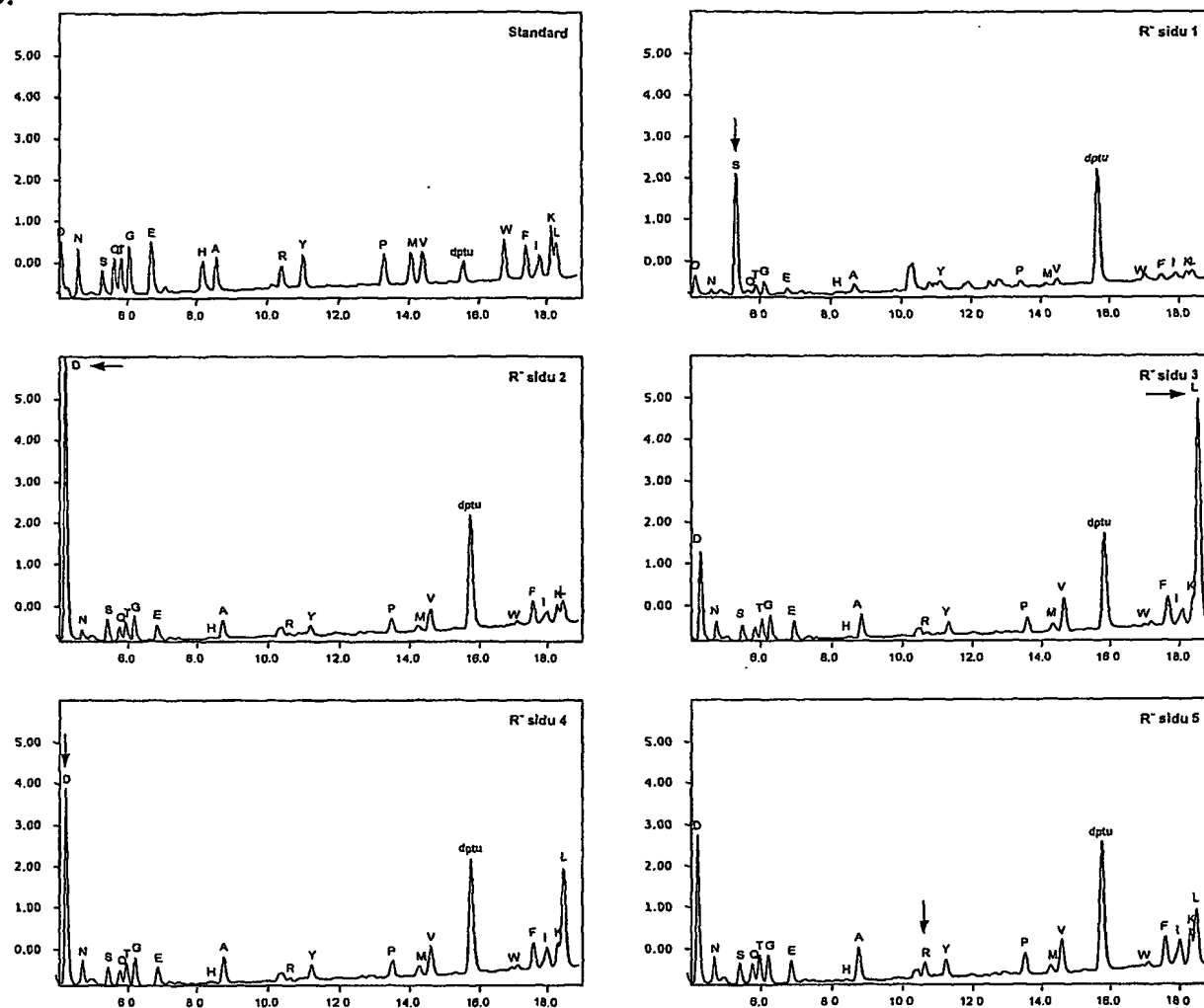
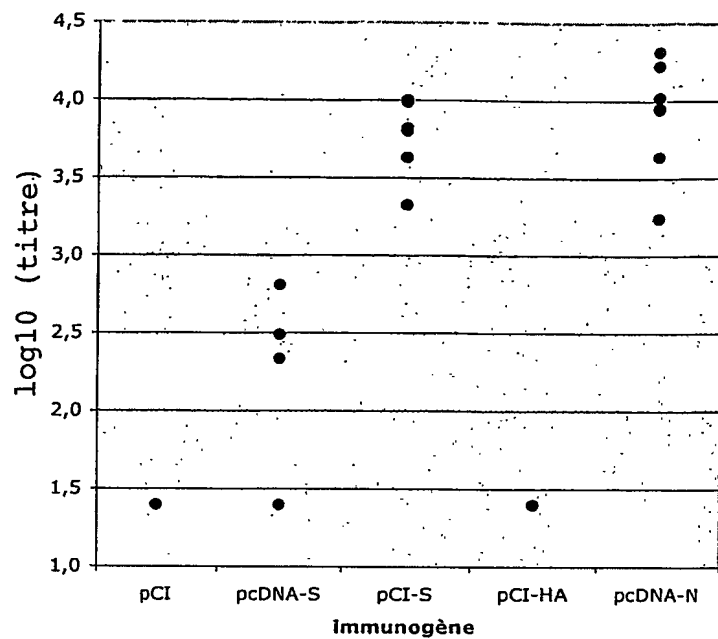


FIGURE 27 D

99/116

A.



B.

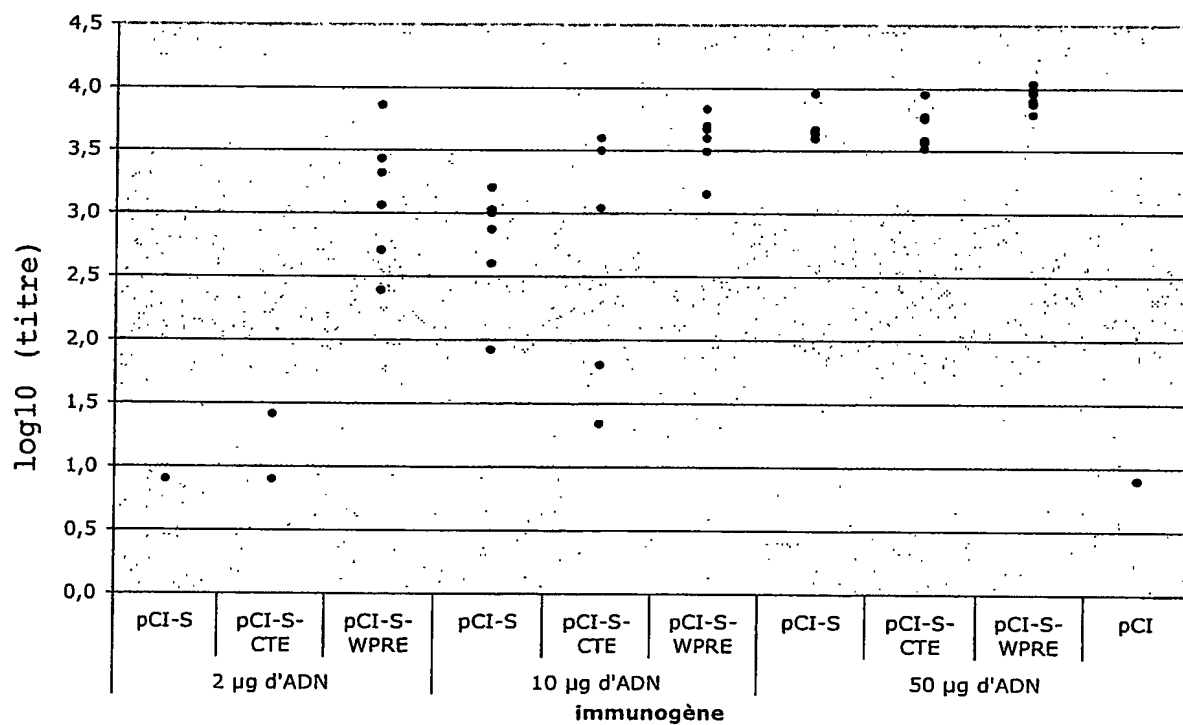
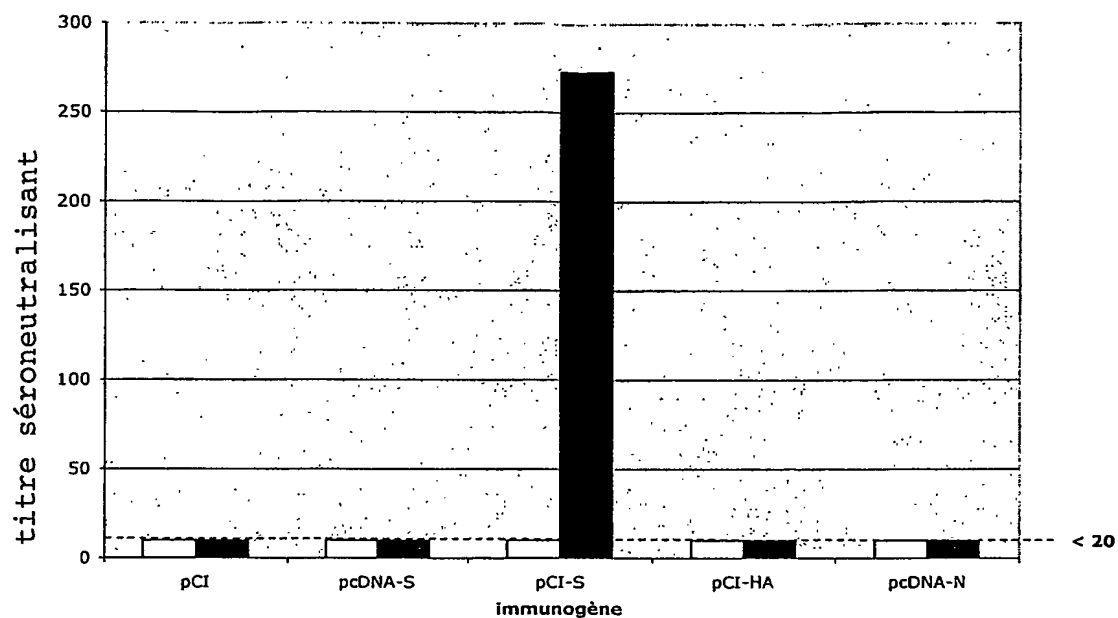


FIGURE 28

A.



B.

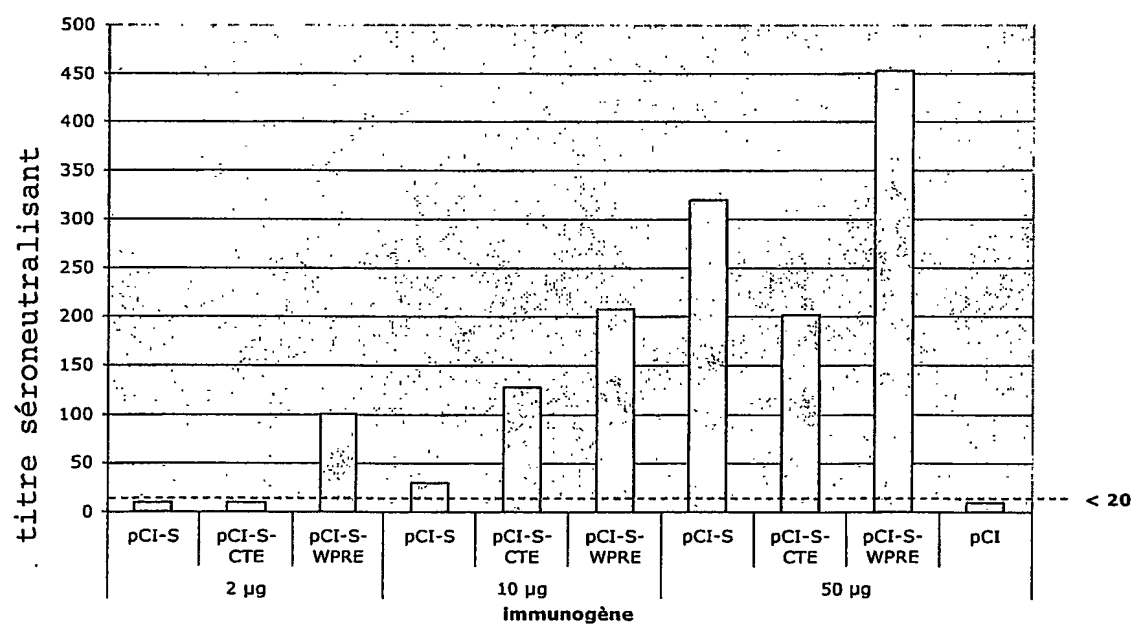


FIGURE 29

101/116

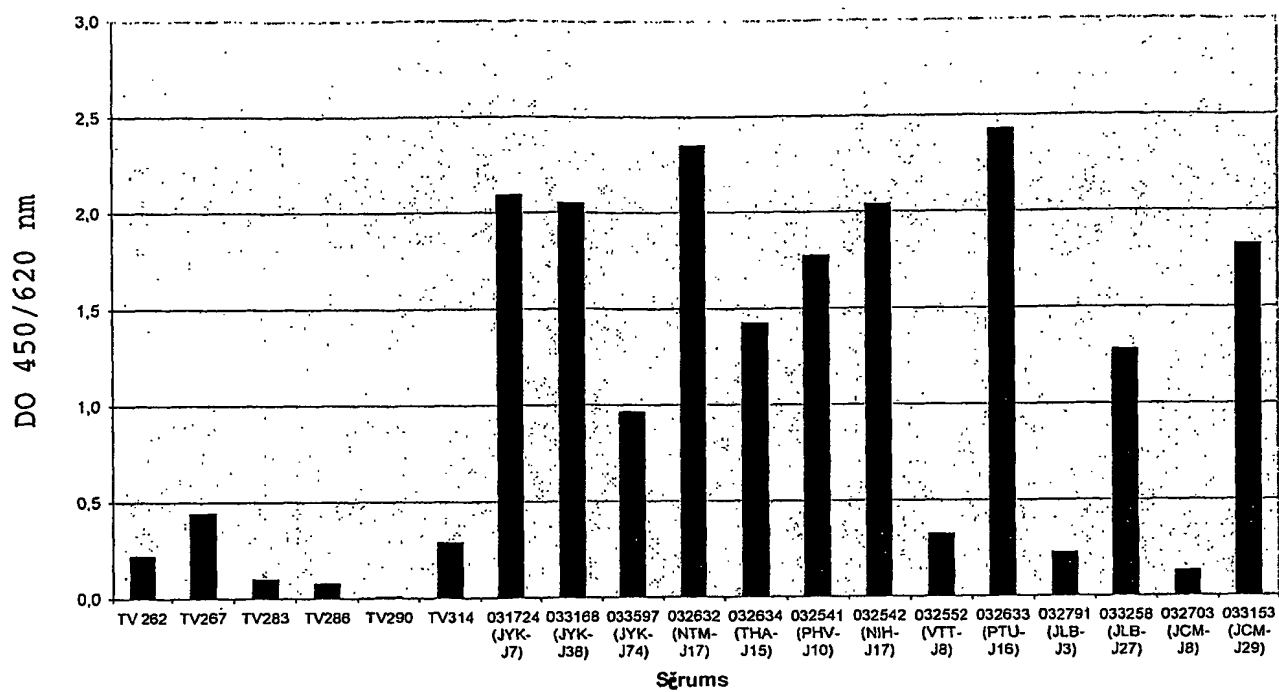


FIGURE 30

102/116

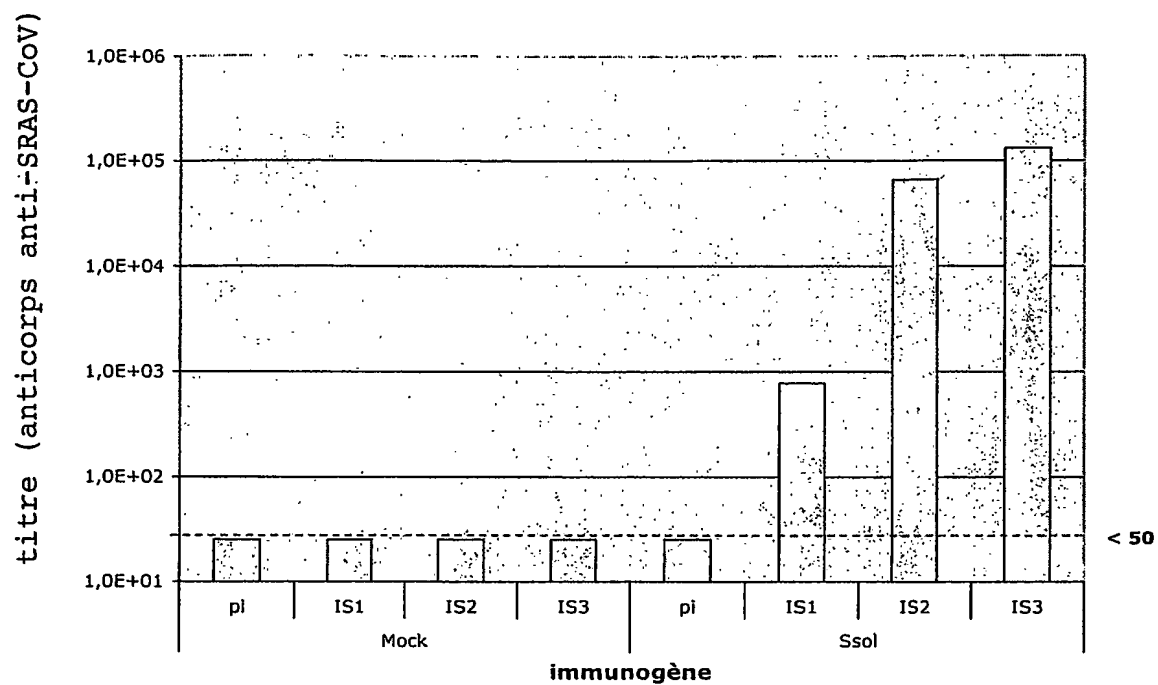


FIGURE 31

103/116

I-3059 1 CTCTTCTGAAAAAGGTAGGCTTATCATTAGAGAAAACAACAGAGTTGTGGTTTCAAGTG
S-040530

I-3059 61 ATATTCTTGTTAACAACATAACGAACATGTTTATTTCTTATTATTTCTTACTCTCACTA
S-040530 1 GG" T" C" C" " " " " C" " C" " " C" " G" C" " G" " C" " G" " C" "

I-3059 121 GTGGTAGTGACCTTGACCGGTGCACCACTTTTGATGATGTTCAAGCTCCTAATTACACTC
S-040530 44 " C" " C" " C" " " " " G" " " " " " " " " " " C" " C" " C" " C" " G" " G" " C" " C" " C" " " " " C" "

I-3059 181 AACATACTT CATCTATGAGGGGGGTTTACTATCCTGATGAAATTTTATAGATCAGACACT
S-040530 104 " G" " C" " C" " CAG" " G" " _ " " " C" " " " C" " G" " " " " C" " C" " C" " G" " C" " " C" " GAGC" " " " C

I-3059 240 CTTTATTTAACTCAGGATTTATTTCTTCCATTTTATTCTAATGTTACAGGGTTTCATACT
S-040530 163 " " G" " CC" G" " C" " " " " CC" G" " C" " " " C" " C" " C" " CAGC" " C" " G" " C" " C" " C" " C" " C" " C" "

I-3059 300 ATTAATCATACGTTTGGCAACCCTGTCATACCTTTTAAGGATGGTATTTATTTTGCTGCC
S-040530 223 " " C" " C" " C" " C" " C" " " " " " C" " G" " C" " C" " C" " " " C" " C" " C" " C" " C" " C" " C" " "

I-3059 360 ACAGAGAAATCAAATGTTGTCCGTGGTGGGTTTTGGTCTACCATGAACAACAAGTCA
S-040530 283 " " C" " " " " GAGC" " C" " C" " G" " G" " G" " C" " " " " G" " C" " CAGC" " " " " " " " " " " AGC

I-3059 420 CAGTCGGTGATTATTATTAACAATTCTACTAATGTTGTTATACGAGCATGTAACTTTGAA
S-040530 343 " " " AGC" " " " " C" " C" " C" " " " " CAGC" " C" " C" " G" " G" " C" " G" " C" " C" " " " " C" " G

I-3059 480 TTGTGTGACAACCCTTTCTTTGCTGTTTCTAAACCCATGGGTACACAGACACATACTATG
S-040530 403 C" " " " C" " " " " " C" " " " " C" " C" " G" " C" " " " " " " C" " C" " " " C" " C" " " C" " C" " "

I-3059 540 ATATTCGATAATGCATTTAATTGCACCTTTCGAGTACATATCTGATGCCTTTTCGCTTGAT
S-040530 463 " " C" " " " " C" " C" " C" " C" " C" " C" " " " " " C" " " " " " " " " " CAGC" " C" " " " CAGC" " G" " C

I-3059 600 GTTTCAGAAAAGTCAGGTAATTTTAAACACTTACGAGAGTTTGTGTTTAAAAATAAGAT
S-040530 523 " " GAGC" " G" " " AGC" " C" " C" " C" " G" " C" " G" " G" " " " C" " " " " C" " G" " C" " G" " C" "

I-3059 660 GGGTTTCTCTATGTTTATAAGGGCTATCAACCTATAGATGTAGTTCGTGATCTACCTTCT
S-040530 583 " " C" " C" " G" " C" " G" " C" " " " " " C" " G" " C" " C" " C" " G" " GA" A" C" G" " CAGC

I-3059 720 GGTTTTAAACACTTTGAAACCTATTTTAAAGTTGCCTCTTGGTATTAAACATTACAAATTTT
S-040530 643 " " C" " C" " " " " CC" " " " " G" " C" " C" " C" " C" " C" " C" " G" " C" " C" " " " " C" " C" " C" " C" "

I-3059 780 AGAGCCATTCTTACAGCCTTTTACCTGCTCAAGACATTGGGGCACGTGAGCTGCAGCC
S-040530 703 C" G" " " " C" " G" " C" " " " " " AGC" " " " " C" " G" " " " C" " " " " " CAGC" " C" " C" " "

I-3059 840 TATTTTGTGGCTATTTAAAGCCAACTACATTTATGCTCAAGTATGATGAAAATGGTACA
S-040530 763 " " C" " C" " G" " " " " CC" G" " " " T" " C" " C" " C" " " " " G" " " " C" " C" " G" " C" " C" " C" "

I-3059 900 ATCACAGATGCTGTTGATTGTTCTCAAAATCCACTTGCTGAACTCAAATGCTCTGTTAAG
S-040530 823 " " " " C" " C" " C" " G" " C" " CAGC" " G" " C" " C" " G" " C" " G" " G" " G" " AGC" " G" " "

I-3059 960 AGCTTTGAGATTGACAAAGGAATTTACCAGACCTCTAATTTAGGGTTGTTCCCTCAGGA
S-040530 883 " " " " C" " " " C" " " " " G" " C" " C" " " " " " " AGC" " C" " " " A" G" G" TAGC" C

I-3059 1020 GATGTTGTGAGATTCCTAATATTACAACTTGTGTCCTTTTGGAGAGGTTTAAATGCT
S-040530 943 " " " " G" " C" G" " " " " C" " " " " C" " C" " C" " C" " C" " C" " C" " A" G" C" " C" " C" "

I-3059 1080 ACTAAATTCCTTCTGTCTATGCATGGGAGAGAAAAAAATTTCTAATTGTGTTGCTGAT
S-040530 1003 " C" " G" " " " CAGC" " G" " C" " C" " " " " C" G" " G" " G" " CAGC" " C" " C" " G" " C" " C

I-3059 1140 TACTCTGTGCTCTACAACCTCAACATTTTTTCAACCTTAAAGTGTATGGCGTTTCTGCC
S-040530 1063 " " " AGC" " " " " G" " " " " " C" " C" " C" " CAGC" " " " C" " " " " " C" " " " " GAGC" " "

I-3059 1200 ACTAAGTTGAATGATCTTTGCTTCTCCAATGTCTATGCAGATTCTTTTGTAGTCAAGGGA
S-040530 1123 " C" " C" " C" " C" " C" " G" " " " " AG" " C" " G" " C" " C" " CAGC" " C" " G" " G" " " " C

I-3059 1260 GATGATGTAAGACAAATAGCGCCAGGACAAACTGGTGTATTGCTGATTATAATTATAAA
S-040530 1183 " C" " C" " G" " " " " G" " C" " C" " C" " T" " C" " G" " C" " C" " G" " C" " C" " C" " C" " C" " C" " G

FIGURE 32.1

I-3059	1320	TGCCAGATGATTTCATGGGTGTCTCCTTGCTTGAATACTAGGAACATTGATGCTACT
S-040530	1243	C""""C"C"C"C""""""C"G"C"G"C""""C"CC""""""C"C"C"C"A
I-3059	1380	TCAACTGGTAATTATAAATTATAAATATAGGTATCTTAGACATGGCAAAGCTTAGGCCCTTT
S-040530	1303	AGC"C"C"C"C"C"C""""C"G"C"CC"C"C"GCG"C""""""GC""""""C
I-3059	1440	GAGAGAGACATATCTAATGTGCCTTTCTCCCCTGATGGCAAACCTTGCACCCACCTGCT
S-040530	1363	""C"G""""C"C"C"C""""C""AG""C"C""""G"C""""""C""""C
I-3059	1500	CTTAATGTATTATTGGCCATTAAATGATTATGGTTTTTACACCACTACTGGCATTGGCTAC
S-040530	1423	"G"C"C"C"C""""CC"G"C"C"C"C"C""""""C"C""""C""""T
I-3059	1560	CAACCTTACAGAGTTGTAGTACTTTCTTTGAACTTTTAAATGCACCGGCCACGGTTTTGT
S-040530	1483	""G"C""""""G"G"G"G"AGC"C"G"GCG"G"C"C"T""""C"G"C
I-3059	1620	GGACCAAAATTATCCACTGACCTTATTAAGAACCAGTGTGTCAATTTTAATTTAATGGA
S-040530	1543	""C""C"GC"GAG""C""""G"C""""""C"G"C""C"C""C"C"
I-3059	1680	CTCACTGGTACTGGTGTGTTAACTCCTTCTTCA_AAGAGATTTCACCATTTCACAAT
S-040530	1603	""G"C"C"C"C""C"G"C""_AG"GC""C"C"C"G"C"C"G"G"
I-3059	1738	TTGGCCGTGATGTCTCTGATTTCACTGATTCCGTTTCGAGATCCTAAAACATCTGAAATAT
S-040530	1661	"C""""G""""GAGC""C""""C"CAG""G"G"C"C"G"CAGC"G"CC
I-3059	1798	TAGACATTTACCTTGTCTCTTTGGGGGTGTAAGTGTAATTACACCTGGAACAAATGCTT
S-040530	1721	"G""""CAGC""C""AGC"C"C"C"C"GTCC"G"C"C"C"C"C"CA
I-3059	1858	_CATCTGAAGTTGCTGTTCTATATCAAGATGTTAACTGCACTGATGTTTCTACAGCAATC
S-040530	1781	G"G"_ ""G"C"G"G"C"G"C"G"C""G""""C"C"GAGC""C"C""
I-3059	1917	CATGCAGATCAACTCACACCAGCTTGGCGCATATATCTACTGAAACAATGTATTCCAG
S-040530	1840	""C""C"C"G"G"G"C"C"C""""G"C"CAGC""C"G""""C"G""""
I-3059	1977	ACTCAAGCAGGCTGTCTTATAGGAGCTGAGCATGTCGACACTTCTTATGAGTGCGACATT
S-040530	1900	""C"G"C""""C"G"C"C"C""""C"G""""CAGC""C""""C
I-3059	2037	CCTATTGGAGCTGGCATTGTGCTAGTTACCATACAGTTTCTTTATTACGTAGTACTAGC
S-040530	1960	""C""""C"C""A""C"C"C"C""""C""GAGCC"GC"G"G"C"C""
I-3059	2097	CAAAAATCTATTGTGGCTTATACTATGCTTTAGGTGCTGATAGTTCAATTGCTTACTCT
S-040530	2020	""G"G"C"C""""C"C"C""AGCC"G"C"C"C"CAGC"C"C""AGC
I-3059	2157	AATAACACCATTGCTATACCTACTAACTTTCAATTAGCATTACTACAGAAGTAATGCCT
S-040530	2080	""C""""""C"C"C"C""C""""CAGC"CTC""C"C"C""""G""""C
I-3059	2217	GTTTCTATGGCTAAAACCTCCGTAGATTGTAATATGTACATCTGCGGAGATTCTACTGAA
S-040530	2140	""GAGC""""C"G"AG""G""C""C""""C""""C""CAGC"C"G
I-3059	2277	TGTGCTAATTTGCTTCTCCAATATGGTAGCTTTTGACACAACATAAATCGTGCACTCTCA
S-040530	2200	""C"C"CC""G"G"G"C"C""""C""""C"G"G"C"G"C"GAGC
I-3059	2337	GGTATTGCTGCTGAACAGGATCGCAACACACGTGAAGTGTTGCTCAAGTCAACAAATG
S-040530	2260	""C"C"C"C"G""G""C"G""""CA"A""""""C""""G"C"G""
I-3059	2397	TACAAAACCCCAACTTTGAAATATTTTGGTGGTTTTAATTTTTCACAAATATTACCTGAC
S-040530	2320	""T""G""""C"CC""G"C"C"G"C"C"C"C"C""T""G"CC"G"C""
I-3059	2457	CCTCTAAAGCCAACTAAGAGGTCTTTTATTGAGGACTTGCTCTTTAATAAGGTGCACTC
S-040530	2380	""G""""C"C"C"C"C"C"C""C""C""""C""""G"C"C"A""""C"G
I-3059	2517	GCTGATGCTGGCTTCATGAAGCAATATGGCGAATGCCTAGGTGATATTAATGCTAGAGAT
S-040530	2440	""C""C"C""""T""""""G"C""""G""""G"C"C"C"C"C"CC"G"C
I-3059	2577	CTCATTTGTGCGCAGAAGTTCAATGGGCTTACAGTGTTGCCACCTTGCTCACTGATGAT
S-040530	2500	"G"C"C"C""C""""""T""C""""G"C"C"C""C"C""C""G"C"C"C
I-3059	2637	ATGATTGCTGCCTACACTGCTGCTCTAGTTAGTGGTACTGCCACTGCTGGATGGACATTT
S-040530	2560	""""C""""T""A""C"C"G"C"C"C"C""""C"C"C""""C"C"

FIGURE 32.2

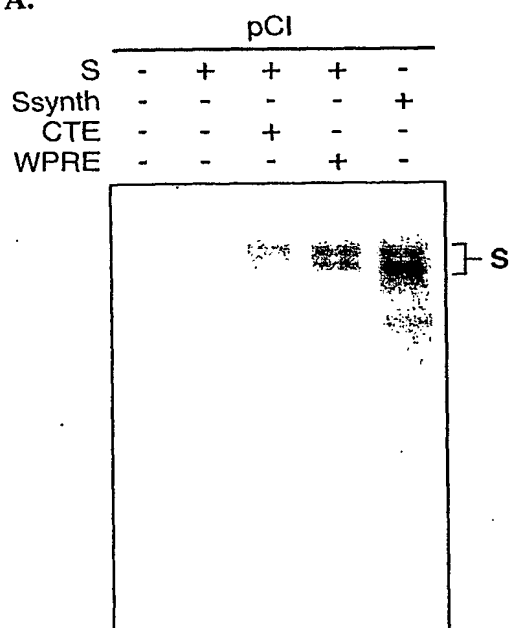
105/116

I-3059	2697	GGTGGCTGGCGCTGCTCTTTCAAATACCTTTTGCTATGCAAATGGCATATAGGTTCAATGGC
S-040530	2620	"A"C"A"C"C"G"G"C"C"C""G""C""CC""C"""
I-3059	2757	ATTGGAGTTACCCAAAATGTTCCTATGAGAACCAAAAACAATCGCCAACCAATTTAAC
S-040530	2680	"C"C"G""G"C"G"G"C""G""G""G""G""G""G""G""C"""
I-3059	2817	AAGGCGATTAGTCAAATTCAAGAATCACTTACAACAACATCAACTGCATTGGGCAAGCTG
S-040530	2740	"C"C"C"C"G"C"G"GAGC"G"C""CAGC"C"CC""C"""
I-3059	2877	CAAGACGTTGTTAACCAGAATGCTCAAGCATTAACACACACTTGTTAAACAACCTAGCTCT
S-040530	2800	"G""G"G"G""C"C"G"CC"G""C"G"G"G"G"G""AGC
I-3059	2937	AATTTTGGTGCAATTTCAAGTGTGCTAAATGATATCCTTTTCGCGACTTGATAAAGTCGAG
S-040530	2860	"C"C"C"C"CAGCTC""G"C"C""GAGCA"G"G"C""G"""
I-3059	2997	GCGGAGGTACAAATTGACAGGCTAATTACAGGCAGACTTCAAAGCCTTCAAACCTATGTA
S-040530	2920	"C"A"G"G"C""C""G"C"C"AC"C"G"GTC""G"G""C"G"
I-3059	3057	ACACAACAATAATCAGGGCTGCTGAAATCAGGGCTTCTGCTAATCTTGCTGCTACTAAA
S-040530	2980	"C"G"G"G""A"C"C"G"C""CAGC"C""G"C"C"C"G"
I-3059	3117	ATGTCTGAGTGTGTTCTTGGACAATCAAAAAGAGTTGACTTTTGTGGAAGGGCTACCAC
S-040530	3040	"AGC""C"G"G"C"GAGC"G""G""C"C"C""T""
I-3059	3177	CTTATGTCCTTCCCACAAGCAGCCCCGCATGGTGTGTCTTCTACATGTCACGTATGTG
S-040530	3100	"G""AG""C"G"C""C"C"G"G""G"C"G"C"C"""
I-3059	3237	CCATCCCAGGAGAGGAACTTCACCACAGCGCCAGCAATTTGTTCATGAAGGCCAAAGCATAC
S-040530	3160	"TAG""C""C""C""C""C""C""C""G""G"C"""
I-3059	3297	TTCCTCGTGAAGGTGTTTTTGTGTTTAAATGGCACTTCTTGTTTATTACACAGAGGAAC
S-040530	3220	"C"G"G"C"G"C""C""C""CAGC""C"C"C"C"C"""
I-3059	3357	TTCTTTTCTCCACAAATAATTACTACAGACAATACATTTGTCTCAGGAAATTGTGATGTC
S-040530	3280	"CAGC"C"G"C"C"C""C""C""G"C"C"C""G
I-3059	3417	GTTATTGGCATCATTAACAACACAGTTTATGATCCTCTGCAACCTGAGCTTGACTCATTC
S-040530	3340	"G"C""C""C""C""G"C"C"C""G"C""G""AGC""
I-3059	3477	AAAGAAGAGCTGGACAAGTACTTCAAAAATCATACATCACAGATGTTGATCTTGCGGAC
S-040530	3400	"G"G""A""G"C"C"C"C"C"C"G"C"G""T
I-3059	3537	ATTTCAGGCATTAACGCTTCTGTGCTCAACATTCAAAAAGAAATTGACCGCCTCAATGAG
S-040530	3460	"CAGC""C""C"G"G""C"G"G"G"C""A"A"G"C"A
I-3059	3597	GTCGCTAAAAATTTAAATGAATCACTCATTGACCTTCAAGAATTGGGAAAATATGAGCAA
S-040530	3520	"G"C"G"CC"G"C"GAGC"G"C""G"G"GC""C"G"C""G
I-3059	3657	TATATTAAATGGCCTTGGTATGTTTGGCTCGGCTTCATTGCTGGACTAATTGCCATCGTC
S-040530	3580	"C"C"G""C""C"G""G""C""C""G""C""G""G
I-3059	3717	ATGTTTACAATCTTGCTTTGTTGCATGACTAGTTGTTGCAGTTGCCTCAAGGGTGCATGC
S-040530	3640	"G"C""G"C""G"C""C""C""C""T""C""G"A"C"C"
I-3059	3777	TCTTGTGGTTCTTGCTGCAAGTTTGATGAGGATGACTCTGAGCCAGTTCTCAAGGGTGTG
S-040530	3700	AGC""CAGC""C""C""C""AGC""C"G"G""C"G
I-3059	3837	AAATTACATTACACATAAACGAACCTTATGGATTGTTTATGAGATTTTTTACTCTTGAT
S-040530	3760	"GC"G"C""C"G"T_""CGA"
I-3059	3897	CAATTACTGCACAGCCAGTAAAAATTGACAATGCTTCTCCTGCAAGT
S-040530		

FIGURE 32.3

106/116

A.



B.

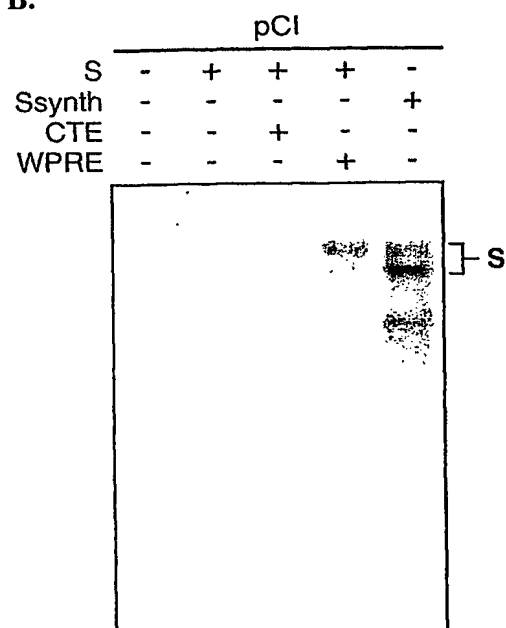
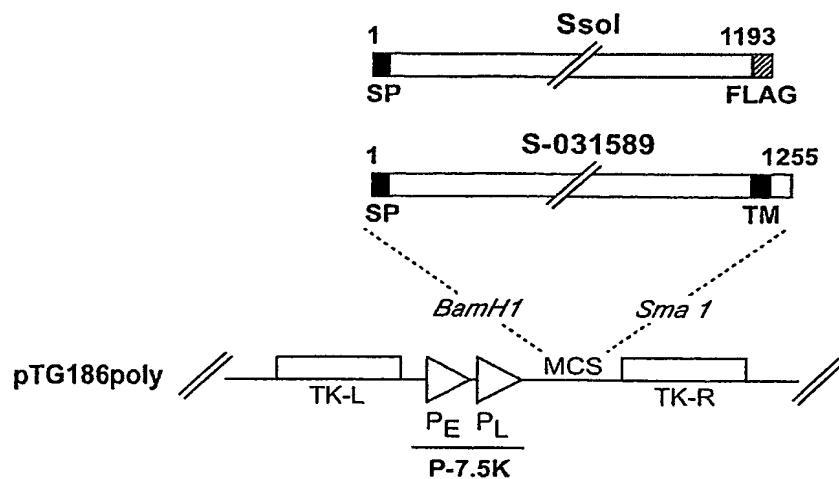


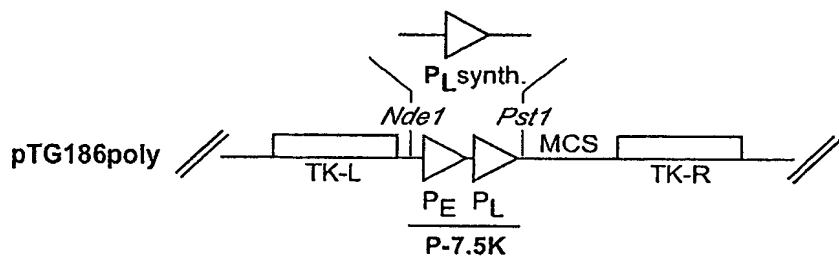
FIGURE 33

107/116

A.



B.



C.

CATATG AGC [T]₂₀GGCATATAAATA GACTC GGCGCGCC AT CTGCAG
*Nde*I promoteur 480 *Asc*I *Pst*I

FIGURE 34 A-C

108/116

D.

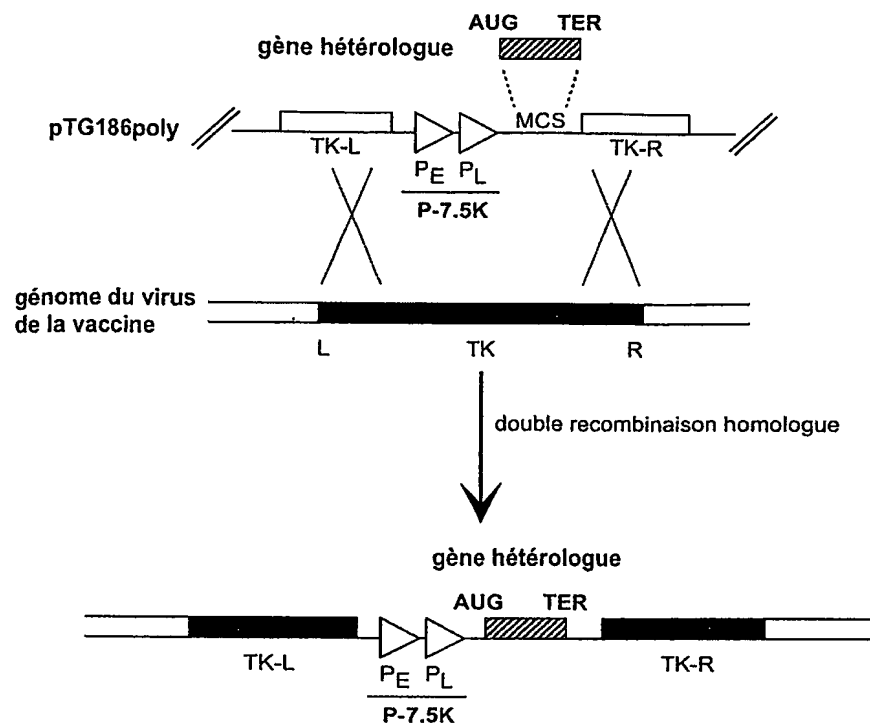
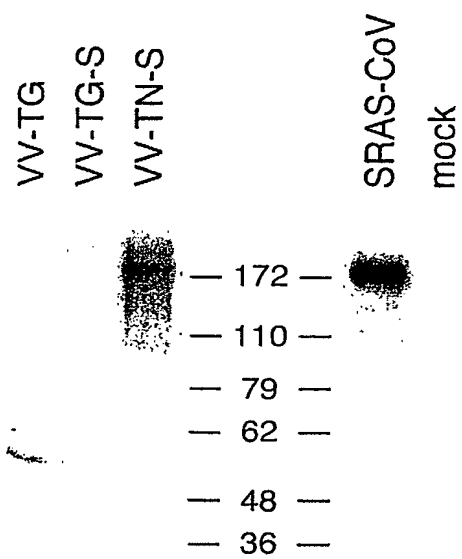


FIGURE 34 D

109/116

A.



B.

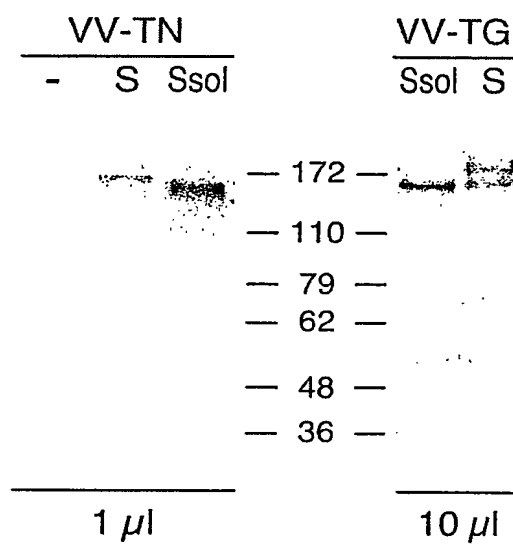
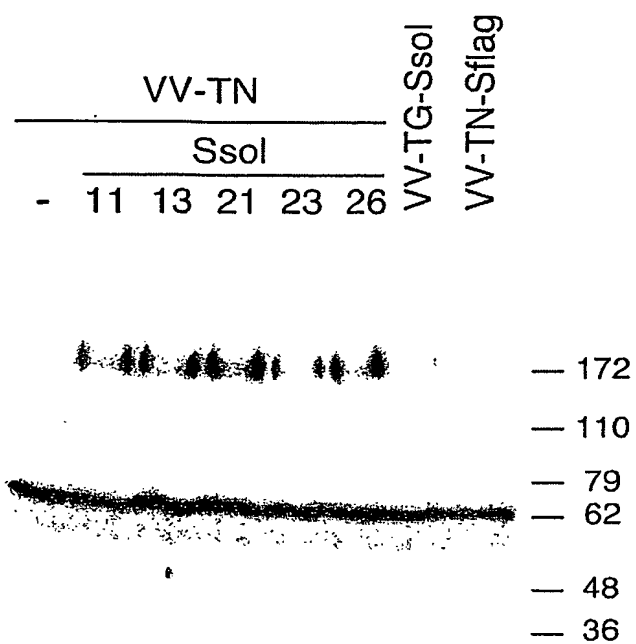


FIGURE 35

110/116

A.



B.

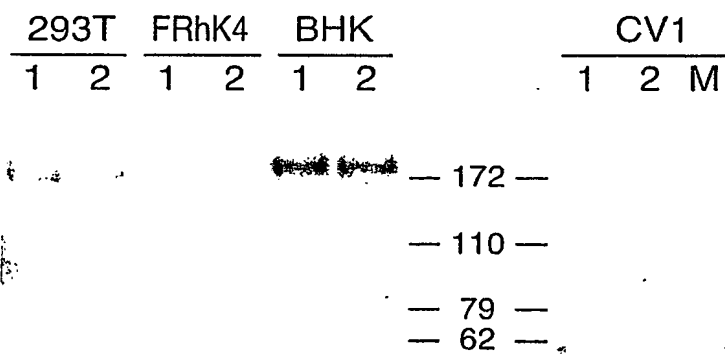


FIGURE 36

111/116

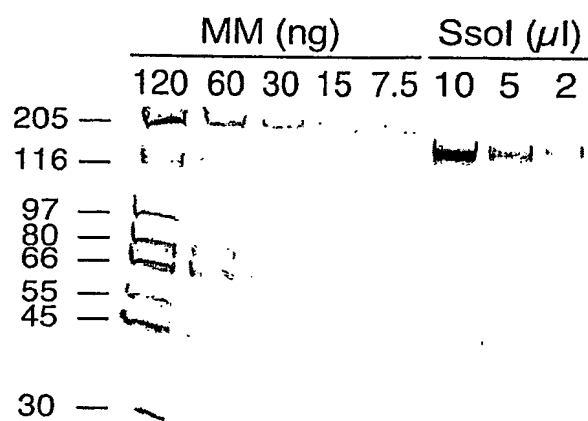


FIGURE 37

112/116

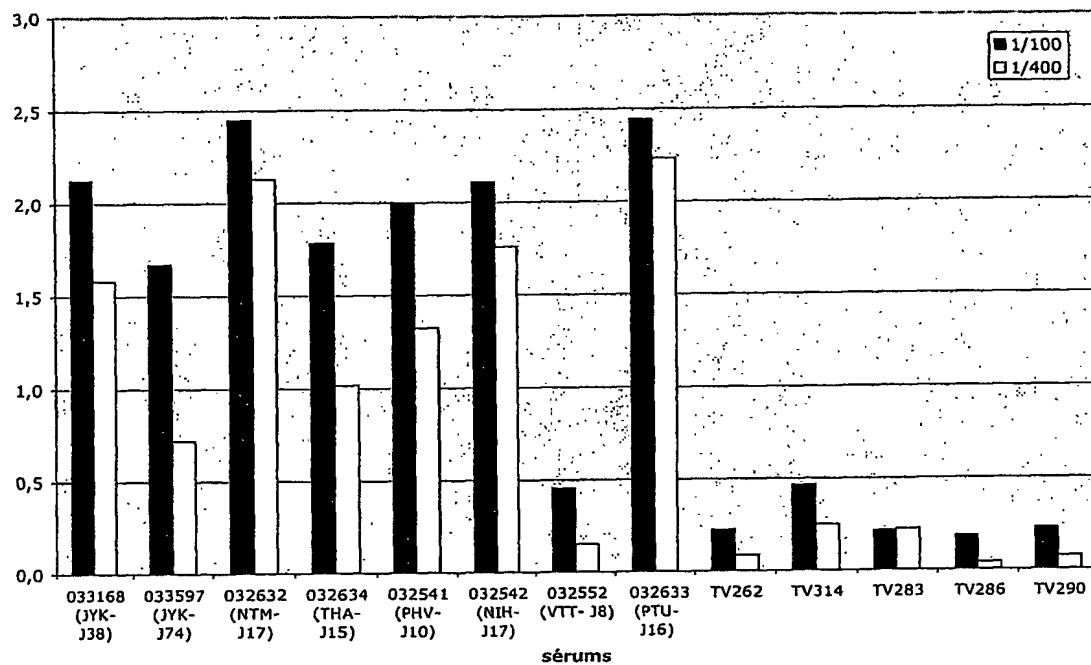
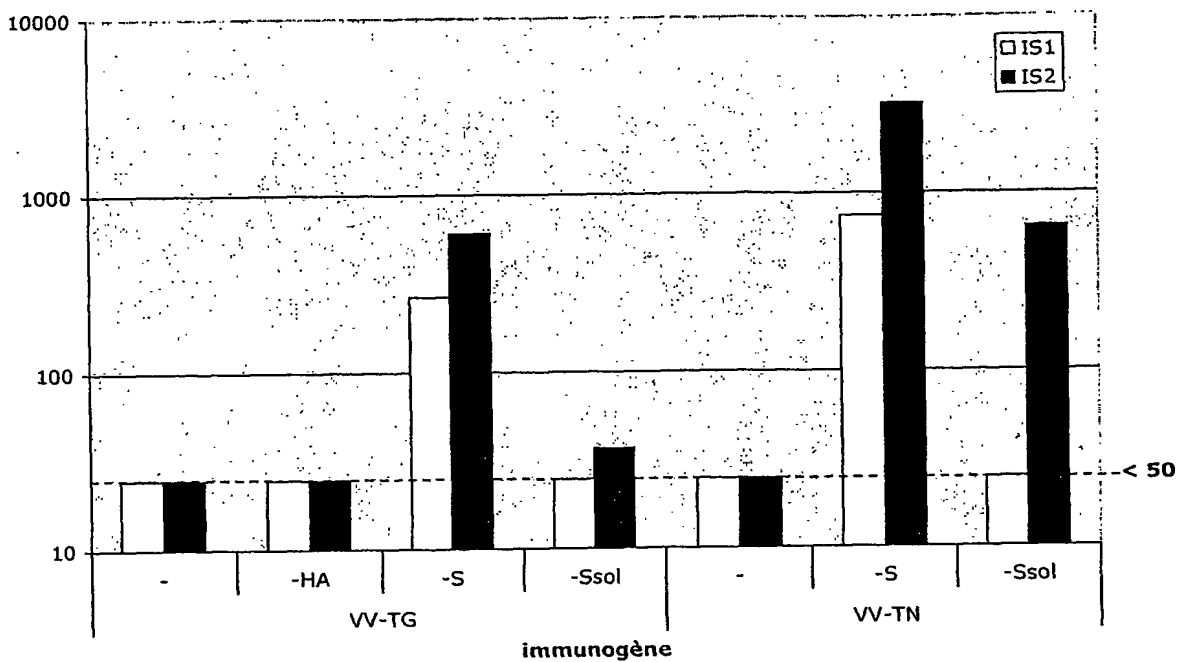


FIGURE 38

113/116

A.



B.

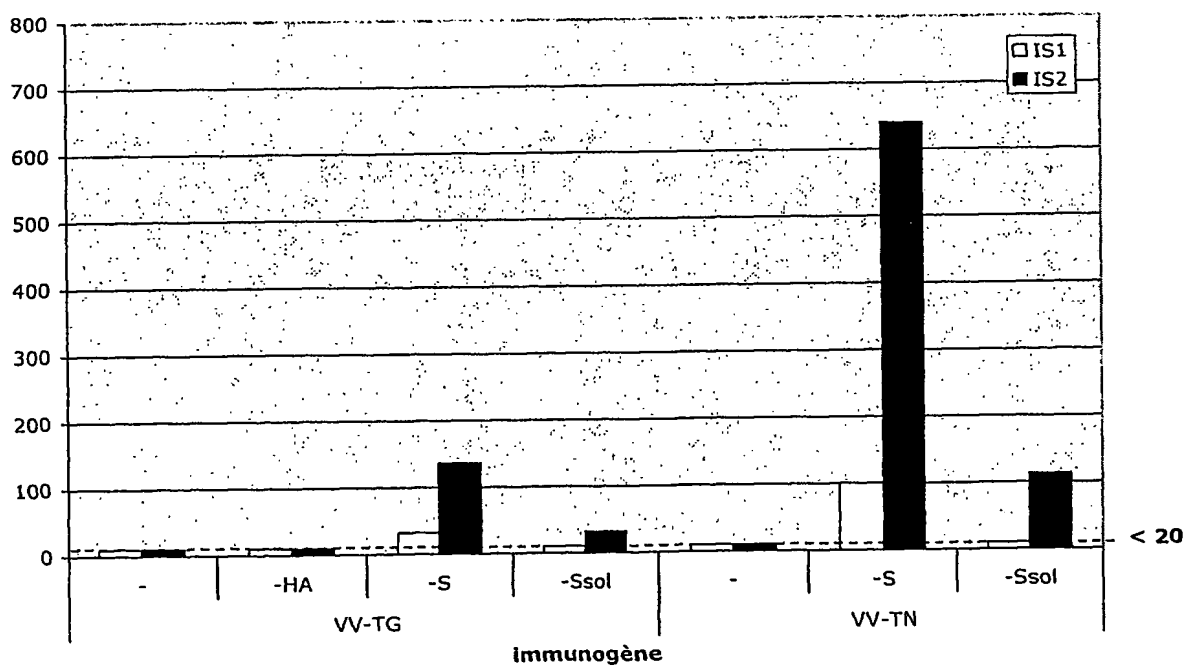


FIGURE 39

114/116

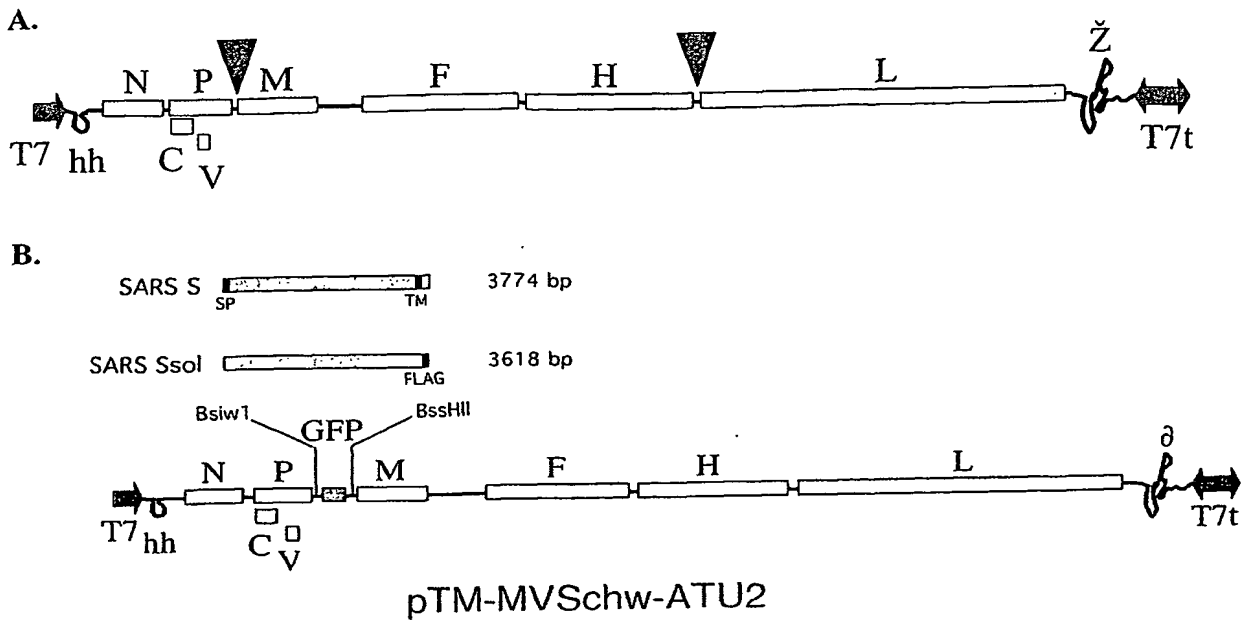


FIGURE 40

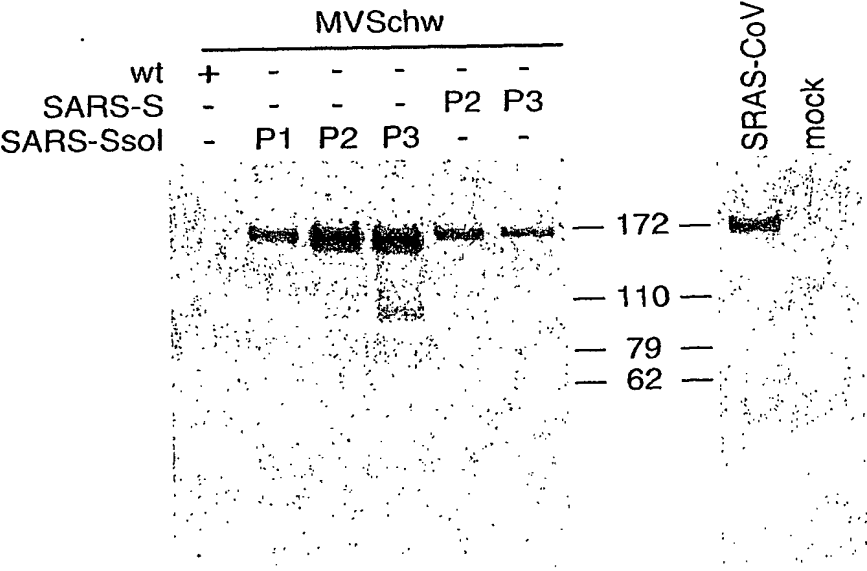


FIGURE 41

115/116

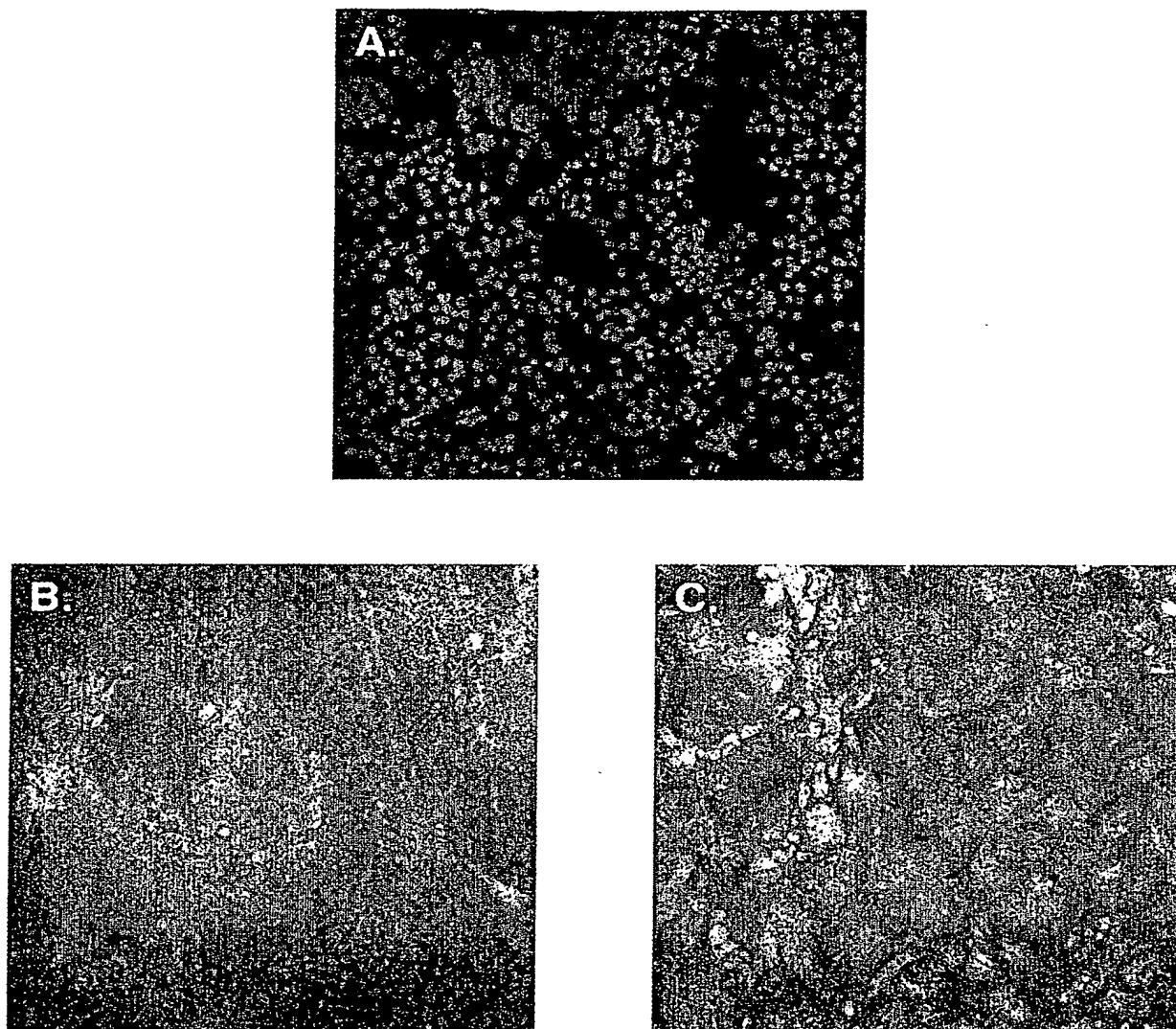


FIGURE 42

116/116

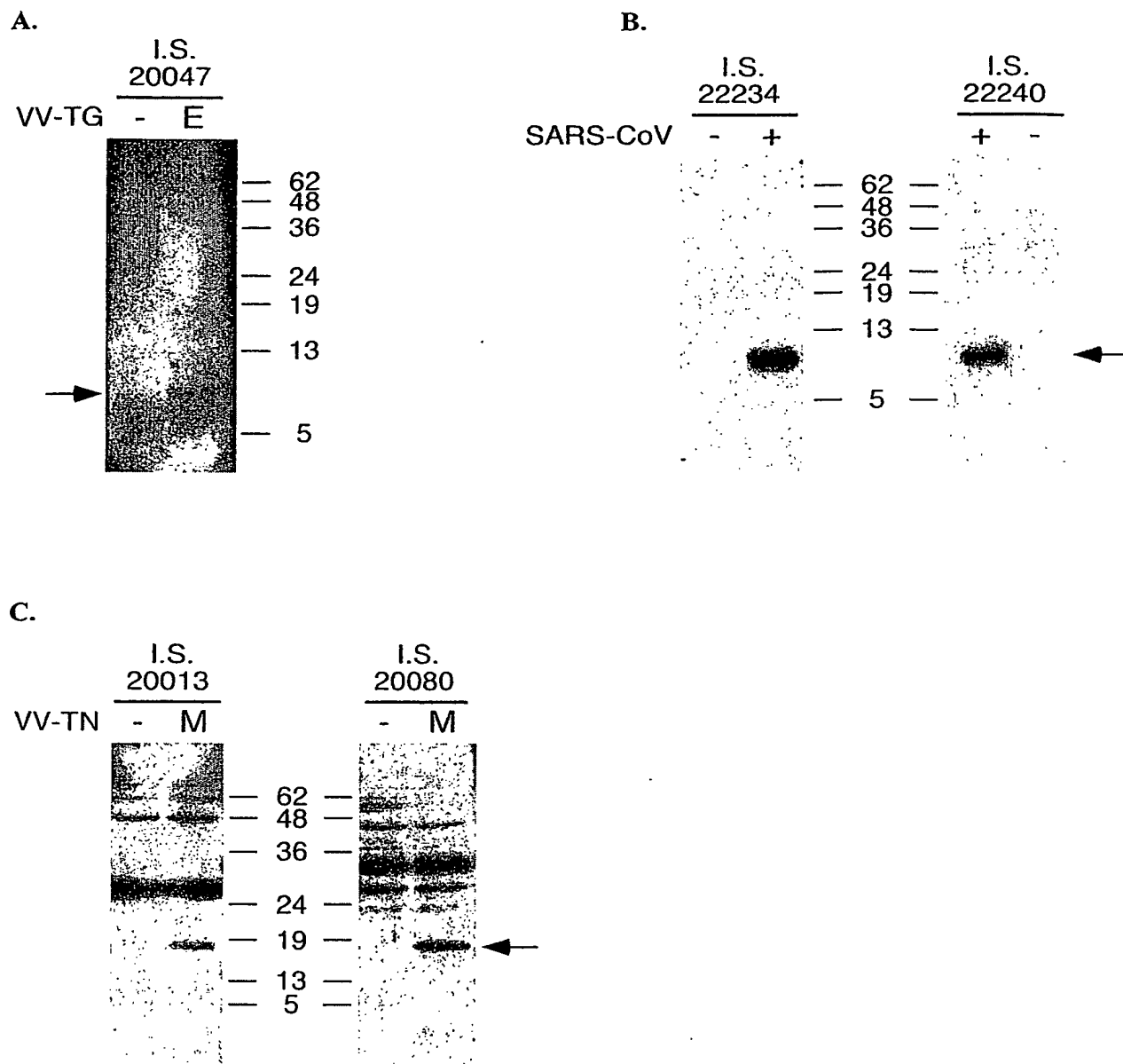


FIGURE 43

S226CAS108.ST25
SEQUENCE LISTING

<110> INSTITUT PASTEUR
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE PARIS 7
VAN DER WERF, Sylvie
ESCRIOU, Nicolas
CRESCENZO-CHAIGNE, Bernadette
MANUGUERRA, Jean-Claude
KUNST, Franck
CALLENDRET, Benoît
BETTON, Jean-Michel
LORIN, Valérie
GERBAUD, Sylvie
BURGUIERE, Ana Maria
AZEBI, Saliha
CHARNEAU, Pierre
TANGY, Frédéric
COMBREDT, Chantal
DELAGNEAU, Jean-François
MARTIN, Monique

<120> NOUVELLE SOUCHE DE CORONAVIRUS ASSOCIE AU SRAS ET SES APPLICATIONS

<130> 226-108ext

<150> FR 0314152
<151> 2003-12-02

<150> FR 0314151
<151> 2003-12-02

<160> 158

S226CAS108.ST25

<170> PatentIn version 3.1

<210> 1

<211> 29746

<212> DNA

<213> CORONAVIRUS

<400> 1

atattagggtt	tttacctacc	caggaaaagc	caaccaacct	cgatctcttg	tagatctggt	60
ctctaaacga	actttaaaat	ctgtgtagct	gtcgctcggc	tgcatgccta	gtgcacctac	120
gcagtataaa	caataataaa	ttttactgtc	gttgacaaga	aacgagtaac	tcgtccctct	180
tctgcagact	gcttacgggt	tcgtccgtgt	tgcagtcgat	catcagcata	cctagggttc	240
gtccgggtgt	gaccgaaagg	taagatggag	agccttggtc	ttggtgtcaa	cgagaaaaca	300
cacgtccaac	tcagtttgcc	tgctcttcag	gttagagacg	tgctagtgcg	tggcttcggg	360
gactctgtgg	aagaggccct	atcggaggca	cgtgaacacc	tcaaaaatgg	cacttggtgt	420
ctagtagagc	tggaaaaagg	cgtagtgccc	cagcttgaac	agccctatgt	gttcattaaa	480
cgttctgatg	ccttaagcac	caatcacggc	cacaaggctc	ttgagctggg	tgcagaaatg	540
gacggcattc	agtacggctg	tagcgggtata	acactgggag	tactcgtgcc	acatgtgggc	600
gaaaccccaa	ttgcataccg	caatgttctt	cttcgtaaga	acggtaataa	gggagccggt	660
ggcatagct	atggcatcga	tctaaagtct	tatgacttag	gtgacgagct	tggcactgat	720
cccattgaag	attatgaaca	aaactggaac	actaagcatg	gcagtgggtg	actccgtgaa	780
ctcactcgtg	agctcaatgg	aggtgcagtc	actcgctatg	tcgacaacaa	tttctgtggc	840
ccagatgggt	accctcttga	ttgcatcaaa	gattttctcg	cacgcgcggg	caagtcaatg	900
tgactctttt	ccgaacaact	tgattacatc	gagtcgaaga	gaggtgtcta	ctgctgccgt	960
gaccatgagc	atgaaattgc	ctggttcact	gagcgctctg	ataagagcta	cgagcaccag	1020
acacccttcg	aaattaagag	tgccaagaaa	tttgacactt	tcaaagggga	atgccc aaag	1080
tttgtgtttc	ctcttaactc	aaaagtcaaa	gtcattcaac	cacgtgttga	aaagaaaaag	1140
actgaggggt	tcatggggcg	tatacgctct	gtgtaccctg	ttgcatctcc	acaggagtgt	1200
aacaatatgc	acttgtctac	cttgatgaaa	tgtaatcatt	gcgatgaagt	ttcatggcag	1260
acgtgcgact	ttctgaaagc	cacttgtgaa	cattgtggca	ctgaaaattt	agttattgaa	1320
ggacctacta	catgtgggta	cctacctact	aatgctgtag	tgaaaatgcc	atgtcctgcc	1380
tgtcaagacc	cagagattgg	acctgagcat	agtggtgcag	attatcacia	ccactcaaac	1440
attgaaactc	gactccgcaa	gggaggtagg	actagatgtt	ttggaggctg	tgtgtttgcc	1500
tatgttggct	gctataataa	gcgtgcctac	tgggttcctc	gtgctagtgc	tgatattggc	1560
tcaggccata	ctggcattac	tggtgacaat	gtggagacct	tgaatgagga	tctccttgag	1620

S226CAS108.ST25

atactgagtc	gtgaacgtgt	taacattaac	attggtggcg	attttcattt	gaatgaagag	1680
gttgccatca	ttttggcatc	tttctctgct	tctacaagtg	cctttattga	cactataaag	1740
agtcttgatt	acaagtcttt	caaaaccatt	gttgagtcct	gcggtacta	taaagttacc	1800
aagggaaagc	ccgtaaaagg	tgcttggaac	attggacaac	agagatcagt	tttaacacca	1860
ctgtgtgggt	ttccctcaca	ggctgctggt	gttatcagat	caatTTTTgc	gcgcacactt	1920
gatgcagcaa	accactcaat	tcctgatttg	caaagagcag	ctgtcaccat	acttgatggt	1980
atttctgaac	agtcattacg	tcttgtcgac	gccatgggtt	atacttcaga	cctgctcacc	2040
aacagtgtca	ttattatggc	atatgtaact	ggtggtcttg	tacaacagac	ttctcagtgg	2100
ttgtctaata	ttttgggcac	tactgttgaa	aaactcaggc	ctatctttga	atggattgag	2160
gcgaaactta	gtgcaggagt	tgaatttctc	aaggatgctt	gggagattct	caaatttctc	2220
attacagggtg	tttttgacat	cgtcaagggt	caaatacagg	ttgcttcaga	taacatcaag	2280
gattgtgtaa	aatgcttcat	tgatgttggt	aacaaggcac	tcgaaatgtg	cattgatcaa	2340
gtcactatcg	ctggcgcaaa	gttgcgatca	ctcaacttag	gtgaagtctt	catcgctcaa	2400
agcaagggac	tttaccgtca	gtgtatacgt	ggcaaggagc	agctgcaact	actcatgcct	2460
cttaaggcac	caaagaagt	aacctttctt	gaagggtgatt	cacatgacac	agtacttacc	2520
tctgaggagg	ttgttctcaa	gaacggtgaa	ctcgaagcac	tcgagacgcc	cgttgatagc	2580
ttcacaaatg	gagctatcgt	tggcacacca	gtctgtgtaa	atggcctcat	gctcttagag	2640
attaaggaca	aagaacaata	ctgcgcattg	tctcctgggt	tactggctac	aaacaatgtc	2700
tttcgcttaa	aaggggggtg	accaattaaa	ggtgtaacct	ttggagaaga	tactgtttgg	2760
gaagttcaag	gttacaagaa	tgtgagaatc	acatttgagc	ttgatgaacg	tgttgacaaa	2820
gtgcttaatg	aaaagtgctc	tgtctacact	gttgaatccg	gtaccgaagt	tactgagttt	2880
gcatgtgttg	tagcagaggc	tgttgatgaa	actttacaac	cagtttctga	tctccttacc	2940
aacatgggta	ttgatcttga	tgagtggagt	gtagctacat	tctacttatt	tgatgatgct	3000
ggtgaagaaa	acttttcatc	acgtatgtat	tgttcctttt	accctccaga	tgaggaagaa	3060
gaggacgatg	cagagtgtga	ggaagaagaa	attgatgaaa	cctgtgaaca	tgagtacggt	3120
acagagggatg	attatcaagg	tctccctctg	gaatttggtg	cctcagctga	aacagttcga	3180
gttgaggaag	aagaagagga	agactggctg	gatgatacta	ctgagcaatc	agagattgag	3240
ccagaaccag	aacctacacc	tgaagaacca	gttaatcagt	ttactgggtta	tttaaaactt	3300
actgacaatg	ttgccattaa	atgtgttgac	atcgtaaggg	aggcacaag	tgctaactct	3360
atggtgattg	taaagtctgc	taacatacac	ctgaaacatg	gtggtgggtg	agcaggtgca	3420
ctcaacaagg	caaccaatgg	tgccatgcaa	aaggagagtg	atgattacat	taagctaaat	3480
ggccctctta	cagtaggagg	gtcttgtttg	ctttctggac	ataatcttgc	taagaagtgt	3540
ctgcatgttg	ttggacctaa	cctaaatgca	ggtgaggaca	tccagcttct	taaggcagca	3600
tatgaaaatt	tcaattcaca	ggacatctta	cttgaccat	tggtgtcagc	aggcatattt	3660

S226CAS108.ST25

ggtgctaaac	cacttcagtc	tttacaagtg	tgcgtgcaga	cggttcgtac	acaggtttat	3720
attgcagtca	atgacaaagc	tctttatgag	caggttgtca	tggattatct	tgataacctg	3780
aagcctagag	tggaagcacc	taaacaagag	gagccaccaa	acacagaaga	ttccaaaact	3840
gaggagaaat	ctgtcgtaca	gaagcctgtc	gatgtgaagc	caaaaattaa	ggcctgcatt	3900
gatgaggtta	ccacaacact	ggaagaaact	aagtttctta	ccaataagtt	actcttgttt	3960
gctgatataca	atggtaagct	ttaccatgat	tctcagaaca	tgcttagagg	tgaagatatg	4020
tctttccttg	agaaggatgc	accttacatg	gtaggatgat	ttatcactag	tggtgatatc	4080
acttggtgtg	taataccctc	caaaaaggct	ggtggcacta	ctgagatgct	ctcaagagct	4140
ttgaagaaag	tgccagttga	tgagtatata	accacgtacc	ctggacaagg	atgtgctggt	4200
tatacacttg	aggaagctaa	gactgctctt	aagaaatgca	aatctgcatt	ttatgtacta	4260
ccttcagaag	cacctaatgc	taaggaagag	attctaggaa	ctgtatcctg	gaatttgaga	4320
gaaatgcttg	ctcatgctga	agagacaaga	aaattaatgc	ctatatgcat	ggatgttaga	4380
gccataatgg	caaccatcca	acgtaagtat	aaaggaatta	aaattcaaga	gggcatcgtt	4440
gactatgggt	tccgattctt	cttttatact	agtaaagagc	ctgtagcttc	tattattacg	4500
aagctgaact	ctctaaatga	gccgcttggtc	acaatgccaa	ttggttatgt	gacacatggt	4560
tttaatcttg	aagaggctgc	gcgctgtatg	cgttctctta	aagctcctgc	cgtagtgtca	4620
gtatcatcac	cagatgctgt	tactacatat	aatggatacc	tcacttcgtc	atcaaagaca	4680
tctgaggagc	actttgtaga	aacagtttct	ttggctggct	cttacagaga	ttggtcctat	4740
tcaggacagc	gtacagagtt	aggtgttgaa	tttcttaagc	gtggtgacaa	aattgtgtac	4800
cacactctgg	agagccccgt	cgagtttcat	cttgacgggtg	aggttctttc	acttgacaaa	4860
ctaaagagtc	tcttatccct	gcgggagggt	aagactataa	aagtgttcac	aactgtggac	4920
aacactaatc	tccacacaca	gcttggtggat	atgtctatga	catatggaca	gcagtttggt	4980
ccaacatact	tggatgggtgc	tgatgttaca	aaaattaaac	ctcatgtaaa	tcatgaggggt	5040
aagactttct	ttgtactacc	tagtgatgac	acactacgta	gtgaagcttt	cgagtactac	5100
catactcttg	atgagagttt	tcttggtagg	tacatgtctg	ctttaaacca	cacaaagaaa	5160
tggaatttc	ctcaagttgg	tggtttaact	tcaattaaat	gggctgataa	caattgttat	5220
ttgtctagt	ttttattagc	acttcaacag	cttgaagtca	aattcaatgc	accagcactt	5280
caagaggctt	attatagagc	ccgtgctggt	gatgctgcta	acttttgtgc	actcactatc	5340
gcttacagta	ataaaactgt	tgccgagctt	ggtgatgtca	gagaaactat	gacccatctt	5400
ctacagcatg	ctaatttgga	atctgcaaag	cgagttctta	atgtgggtgtg	taaacattgt	5460
ggtcagaaaa	ctactacctt	aacgggtgta	gaagctgtga	tgtatatggg	tactctatct	5520
tatgataatc	ttaagacagg	tgtttccatt	ccatgtgtgt	gtggctcgtga	tgctacacaa	5580
tatctagtag	aacaagagtc	ttcttttggt	atgatgtctg	caccacctgc	tgagtataaa	5640
ttacagcaag	gtacattctt	atgtgcgaat	gagtacactg	gtaactatca	gtgtgggtcat	5700

S226CAS108.ST25

tacactcata	taactgctaa	ggagaccctc	tatcgtattg	acggagctca	ccttacaaag	5760
atgtcagagt	acaaaggacc	agtgactgat	gttttctaca	aggaaacatc	ttacactaca	5820
accatcaagc	ctgtgtcgta	taaactcgat	ggagttactt	acacagagat	tgaacaaaaa	5880
ttggatgggt	attataaaaa	ggataatgct	tactatacag	agcagcctat	agaccttgta	5940
ccaactcaac	cattaccaaa	tgcgagtttt	gataatttca	aactcacatg	ttctaacaca	6000
aaatttgctg	atgattttaa	tcaaattgaca	ggcttcacaa	agccagcttc	acgagagcta	6060
tctgtcacat	tcttcccaga	cttgaatggc	gatgtagtgg	ctattgacta	tagacactat	6120
tcagcgagtt	tcaagaaagg	tgctaaatta	ctgcataagc	caattgtttg	gcacattaac	6180
caggctacaa	ccaagacaac	gttcaaacca	aacacttggt	gtttacgttg	tctttggagt	6240
acaaagccag	tagatacttc	aaattcattt	gaagttctgg	cagtagaaga	cacacaagga	6300
atggacaatc	ttgcttgtga	aagtcaacaa	cccacctctg	aagaagtagt	ggaaaaatcct	6360
accatacaga	aggaagtcac	agagtgtgac	gtgaaaacta	ccgaagttgt	aggcaatgtc	6420
atacttaaac	catcagatga	aggtgttaaa	gtaacacaag	agttagggtca	tgaggatcct	6480
atggctgctt	atgtggaaaa	cacaagcatt	accattaaga	aacctaata	gctttcacta	6540
gccttagggt	taaaaacaat	tgccactcat	ggtattgctg	caattaatag	tgttccttgg	6600
agtaaaat	ttggcttatgt	caaaccattc	ttaggacaag	cagcaattac	aacatcaa	6660
tgcgctaaga	gattagcaca	acgtgtgttt	aacaattata	tgcttatgt	gtttacatta	6720
ttgttccaat	tgtgtacttt	tactaaaagt	accaattcta	gaattagagc	ttcactacct	6780
acaactattg	ctaaaaatag	tgttaagagt	gttgctaaat	tatgtttgga	tgccggcatt	6840
aattatgtga	agtcacccaa	atthttctaa	ttgttcacaa	tcgctatgtg	gctattgttg	6900
ttaagtattt	gcttaggttc	tctaactctgt	gtaactgctg	cttttggtgt	actcttatct	6960
aattttggtg	ctccttctta	ttgtaatggc	gttagagaat	tgtatcttaa	ttcgtcta	7020
gttactacta	tggatttctg	tgaaggttct	tttccttgca	gcatttgttt	aagtggatta	7080
gactcccttg	attcttatcc	agctcttgaa	accattcagg	tgacgatttc	atcgta	7140
ctagacttga	caatttttag	tctggccgct	gagtggtttt	tggcatatat	gttggtcaca	7200
aaattctttt	atttatttag	tctttcagct	ataatgcagg	tgttcttttg	ctattttgct	7260
agtcatttca	tcagcaattc	ttggctcatg	tggtttatca	ttagtattgt	acaaatggca	7320
cccgtttctg	caatgggttag	gatgtacatc	ttctttgctt	ctttctacta	catatggaag	7380
agctatgttc	atatcatgga	tggttgcacc	tcttcgactt	gcatgatgtg	ctataagcgc	7440
aatcgtgcca	cacgcgttga	gtgtacaact	attgttaatg	gcatgaagag	atctttctat	7500
gtctatgcaa	atggaggccg	tggcttctgc	aagactcaca	attggaattg	tctcaattgt	7560
gacacatttt	gactggtag	tacattcatt	agtgatgaag	ttgctcgtga	tttgtcactc	7620
cagtttaaaa	gaccaatcaa	ccctactgac	cagtcatcgt	atattgttga	tagtggtgct	7680
gtgaaaaatg	gcgcgcttca	cctctacttt	gacaaggctg	gtcaaaagac	ctatgagaga	7740

S226CAS108.ST25

catccgctct	cccattttgt	caatttagac	aatttgagag	ctaacaacac	taaaggttca	7800
ctgcctatta	atgtcatagt	ttttgatggc	aagtccaaat	gcgacgagtc	tgcttctaag	7860
tctgcttctg	tgtactacag	tcagctgatg	tgccaaccta	ttctgttgct	tgaccaagct	7920
cttgatcag	acgttggaga	tagtactgaa	gtttccgtta	agatgtttga	tgcttatgtc	7980
gacacctttt	cagcaacttt	tagtgttcct	atggaaaaac	ttaaggcact	tgttgctaca	8040
gctcacagcg	agttagcaaa	gggtgtagct	ttagatgggtg	tcctttctac	attcgtgtca	8100
gctgcccagc	aagggtgttg	tgataccgat	gttgacacaa	aggatgttat	tgaatgtctc	8160
aaactttcac	atcactctga	cttagaagtg	acaggtgaca	gttgtaacaa	tttcatgctc	8220
acctataata	aggttgaaaa	catgacgcc	agagatcttg	gcgcatgtat	tgactgtaat	8280
gcaaggcata	tcaatgccc	agtagcaaaa	agtcacaatg	tttcactcat	ctggaatgta	8340
aaagactaca	tgtctttatc	tgaacagctg	cgtaaaca	ttcgtagtgc	tgccaagaag	8400
aacaacatac	cttttagact	aacttggtg	acaactagac	aggttggtcaa	tgtcataact	8460
actaaaatct	cactcaagg	tggttaagatt	gttagtactt	gttttaaact	tatgcttaag	8520
gccacattat	tgtgcgttct	tgctgcattg	gtttgttata	tcgttatgcc	agtacataca	8580
ttgtcaatcc	atgatgggtta	cacaaatgaa	atcattgggt	acaaagccat	tcaggatggt	8640
gtcactcgtg	acatcatttc	tactgatgat	tgttttgcaa	ataaacatgc	tggttttgac	8700
gcatggttta	gccagcgtgg	tggttcatac	aaaaatgaca	aaagctgccc	tgtagtagct	8760
gctatcatta	caagagagat	tggtttcata	gtgcctggct	taccgggtac	tgtgctgaga	8820
gcaatcaatg	gtgacttctt	gcattttcta	cctcgtgttt	ttagtgctgt	tggcaacatt	8880
tgctacacac	cttccaaact	cattgagtat	agtgattttg	ctacctctgc	ttgcgttctt	8940
gctgctgagt	gtacaatttt	taaggatgct	atgggcaaac	ctgtgccata	ttgttatgac	9000
actaatttgc	tagaggggttc	tattttcttat	agtgagcttc	gtccagacac	tcgttatgtg	9060
cttatggatg	gttccatcat	acagtttcct	aacacttacc	tggaggggttc	tgtagagta	9120
gtaacaactt	ttgatgctga	gtactgtaga	catggtacat	gcgaaaggctc	agaagtaggt	9180
atttgccat	ctaccagtg	tagatgggtt	cttaataatg	agcattacag	agctctatca	9240
ggagttttct	gtgggtgttg	tgcatgaat	ctcatagcta	acatctttac	tcctcttggtg	9300
caacctgtgg	gtgctttaga	tgtgtctgct	tcagtagtg	ctggtggtat	tattgccata	9360
ttggtgactt	gtgctgccta	ctactttatg	aaattcagac	gtgttttttg	tgagtacaac	9420
catgttggtg	ctgctaatac	acttttggtt	ttgatgtctt	tcactatact	ctgtctggta	9480
ccagcttaca	gctttctgcc	gggagctctac	tcagtctttt	acttgacttt	gacattctat	9540
ttcaccaatg	atgtttcatt	cttggtcac	cttcaatggt	ttgccatgtt	ttctcctatt	9600
gtgccttttt	ggataacagc	aatctatgta	ttctgtattt	ctctgaagca	ctgccattgg	9660
ttctttaaca	actatcttag	gaaaagagtc	atgtttaatg	gagttacatt	tagtaccttc	9720
gaggaggctg	ctttgtgtac	ctttttgctc	aacaaggaaa	tgtacctaaa	attgcgtagc	9780

S226CAS108.ST25

gagacactgt	tgccacttac	acagtataac	aggatatctg	ctctatataa	caagtacaag	9840
tatttcagtg	gagccttaga	tactaccagc	tatcgtgaag	cagcttgctg	ccacttagca	9900
aaggctctaa	atgacttttag	caactcaggt	gctgatgttc	tctaccaacc	accacagaca	9960
tcaatcactt	ctgctgttct	gcagagtggg	tttaggaaaa	tggcattccc	gtcaggcaaa	10020
gttgaagggg	gcatgggtaca	agtaacctgt	ggaactacaa	ctcttaatgg	attgtgggtg	10080
gatgacacag	tatactgtcc	aagacatgtc	atttgcacag	cagaagacat	gcttaatcct	10140
aactatgaag	atctgctcat	tcgcaaacc	aaccatagct	ttcttggtca	ggctggcaat	10200
gttcaacttc	gtgttattgg	ccatttctatg	caaaattgtc	tgcttaggct	taaagttgat	10260
acttctaacc	ctaagacacc	caagtataaa	tttgtccgta	tccaacctgg	tcaaacattt	10320
tcagttctag	catgctacaa	tgggtcacca	tctgggtgtt	atcagtgtgc	catgagacct	10380
aatcatacca	ttaaagggttc	tttccttaat	ggatcatgtg	gtagtggttg	ttttaacatt	10440
gattatgatt	gcgtgtcttt	ctgctatatg	catcatatgg	agcttccaac	aggagtacac	10500
gctgggtactg	acttagaagg	taaattctat	gggtccattt	ttgacagaca	aactgcacag	10560
gctgcaggta	cagacacaac	cataacatta	aatgttttgg	catggctgta	tgctgctggt	10620
atcaatgggtg	ataggtgggt	tcttaataga	ttcaccacta	ctttgaatga	ctttaacctt	10680
gtggcaatga	agtacaacta	tgaacctttg	acacaagatc	atgttgacat	attgggacct	10740
ctttctgctc	aaacaggaat	tgccgtctta	gatatgtgtg	ctgctttgaa	agagctgctg	10800
cagaatggta	tgaatggctg	tactatcctt	ggtagcacta	ttttagaaga	tgagtttaca	10860
ccatttgatg	ttgttagaca	atgctctggg	gttaccttcc	aaggtaagtt	caagaaaatt	10920
gttaagggca	ctcatcattg	gatgctttta	actttcttga	catcactatt	gattcttggt	10980
caaagtacac	agtggtcact	gtttttcttt	gtttacgaga	atgctttctt	gccattttact	11040
cttgggtatta	tggcaattgc	tgcatgtgct	atgctgcttg	ttaagcataa	gcacgcattc	11100
ttgtgcttgt	ttctgttacc	ttctcttgca	acagttgctt	actttaatat	gggtctacatg	11160
cctgctagct	gggtgatgcg	tatcatgaca	tggcttgaat	tggctgacac	tagcttgtct	11220
ggttataggc	ttaaggattg	tgttatgtat	gcttcagctt	tagttttgct	tattctcatg	11280
acagctcgca	ctgtttatga	tgatgctgct	agacgtgttt	ggacactgat	gaatgtcatt	11340
acacttgttt	acaaagtcta	ctatggtaat	gcttttagatc	aagctatttc	catgtgggcc	11400
ttagttattt	ctgtaacctc	taactattct	gggtgctgta	cgactatcat	gttttttagct	11460
agagctatag	tgtttggtgtg	tgttgagtat	taccatttgt	tatttattac	tggcaacacc	11520
ttacagtgtg	tcatgcttgt	ttattgtttc	ttaggctatt	gttgctgctg	ctactttggc	11580
cttttctgtt	tactcaaccg	ttacttcagg	cttactcttg	gtgtttatga	ctacttgggtc	11640
tctacacaag	aatttaggta	tatgaactcc	caggggcttt	tgccctcctaa	gagtagtatt	11700
gatgctttca	agcttaacat	taagttgttg	ggatttggag	gtaaaccatg	tatcaagggt	11760
gctactgtac	agtctaaaat	gtctgacgta	aagtgcacat	ctgtgggtact	gctctcggtt	11820

S226CAS108.ST25

```

ttcaacaac ttagagtaga gtcatttct aaattgtggg cacaatgtgt acaactccac 11880
aatgatattc ttcttgcaaa agacacaact gaagctttcg agaagatggg ttctcttttg 11940
tctgttttgc tatccatgca ggggtgctgta gacattaata ggttgtgcga ggaaatgctc 12000
gataaccgtg ctactcttca ggctattgct tcagaattta gttctttacc atcatatgcc 12060
gcttatgcca ctgcccagga ggcctatgag caggctgtag ctaatgggtga ttctgaagtc 12120
gttctcaaaa agttaaagaa atctttgaat gtggctaaat ctgagtttga ccgtgatgct 12180
gccatgcaac gcaagttgga aaagatggca gatcaggcta tgacccaaat gtacaaacag 12240
gcaagatctg aggacaagag ggcaaaaagta actagtgtga tgcaaacaaat gctcttcact 12300
atgcttagga agcttgataa tgatgcactt aacaacatta tcaacaatgc gcgtgatggg 12360
tgtgttcac tcaacatcat accattgact acagcagcca aactcatggg tgttgtccct 12420
gattatggta cctacaagaa cacttgtgat ggtaacacct ttacatatgc atctgcactc 12480
tggaataatc agcaagttgt tgatgcggat agcaagattg ttcaacttag tgaaattaac 12540
atggacaatt caccaaattt ggcttggcct cttattgtta cagctctaag agccaactca 12600
gctgttaaac tacagaataa tgaactgagt ccagtagcac tacgacagat gtcctgtgcy 12660
gctggtacca cacaaacagc ttgtactgat gacaatgcac ttgcctacta taacaattcg 12720
aaggagggtg ggtttgtgct ggcattacta tcagaccacc aagatctcaa atgggctaga 12780
ttccctaaga gtgatggtag aggtacaatt tacacagaac tggaaccacc ttgtaggttt 12840
gttacagaca caccaaaagg gcctaaagtg aaatacttgt acttcatcaa aggcttaaac 12900
aacctaata gaggtatggg gctgggcagt ttagctgtga cagtacgtct tcaggctgga 12960
aatgctacag aagtacctgc caattcaact gtgctttcct tctgtgcttt tgcagtagac 13020
cctgctaaag catataagga ttacctagca agtgaggagc aaccaatcac caactgtgtg 13080
aagatgttgt gtacacacac tggtagagga caggcaatta ctgtaacacc agaagctaac 13140
atggaccaag agtccttttg tgggtgcttca tgttgtctgt attgtagatg ccacattgac 13200
catccaaatc ctaaaggatt ctgtgacttg aaaggtaagt acgtccaaat acctaccact 13260
tgtgctaattg acccagtggt ttttacctt agaaacacag tctgtaccgt ctgcggaatg 13320
tggaagggtt atggctgtag ttgtgaccaa ctccgcgaac ccttgatgca gtctgcggat 13380
gcatcaacgt ttttaaaccg gtttgcggtg taagtgcagc ccgtcttaca ccgtgcggca 13440
caggcactag tactgatgtc gtctacaggg cttttgatat ttacaacgaa aaagttgctg 13500
gttttgcaaa gttcctaaaa actaattgct gtcgcttcca ggagaaggat gaggaaggca 13560
atattattaga ctcttacttt gtagttaaga ggcatactat gtctaactac caacatgaag 13620
agactattta taacttggtt aaagattgtc cagcggttgc tgtccatgac tttttcaagt 13680
ttagagtaga tgggtgacatg gtaccacata tatcacgtca gcgtctaact aaatacacia 13740
tggtgtgattt agtctatgct ctacgtcatt ttgatgaggg taattgtgat acattaaaag 13800
aaatactcgt cacatacaat tgctgtgatg atgattattt caataagaag gattggtatg 13860

```

S226CAS108.ST25

acttcgtaga gaatcctgac atcttacgcg tatatgctaa cttaggtgag cgtgtacgcc 13920
 aatcattatt aaagactgta caattctgcg atgctatgcg tgatgcaggc attgtaggcg 13980
 tactgacatt agataatcag gatcttaatg ggaactggta cgatttcggt gatttcgtac 14040
 aagtagcacc aggctgcgga gttcctattg tggattcata ttactcattg ctgatgccca 14100
 tcctcacttt gactagggca ttggctgctg agtcccatat ggatgctgat ctcgcaaaac 14160
 cacttattaa gtgggatttg ctgaaatatg attttacgga agagagactt tgtctcttcg 14220
 accgttattt taaatatttg gaccagacat accatcccaa ttgtattaac tgtttggatg 14280
 atagggtgat ccttcattgt gcaaacttta atgtgttatt ttctactgtg tttccaccta 14340
 caagttttgg accactagta agaaaaatat ttgtagatgg tgttcctttt gttgtttcaa 14400
 ctggatacca ttttcgtgag ttaggagtcg tacataatca ggatgtaaac ttacatagct 14460
 cgcgctctcag tttcaaggaa cttttagtgt atgctgctga tccagctatg catgcagctt 14520
 ctggcaattt attgctagat aaacgcacta catgcttttc agtagctgca ctaacaaca 14580
 atgttgcttt tcaaactgtc aaaccggta attttaataa agacttttat gactttgctg 14640
 tgtctaaagg tttctttaag gaaggaagt ctgttgaact aaaacacttc ttctttgctc 14700
 aggatggcaa cgctgctatc agtgattatg actattatcg ttataatctg ccaacaatgt 14760
 gtgatatcag acaactccta ttcgtagttg aagttgttga taaatacttt gattgttacg 14820
 atgggtggctg tattaatgcc aaccaagtaa tcgttaacaa tctggataaa tcagctgggt 14880
 tcccatttaa taaatggggt aaggctagac tttattatga ctcaatgagt tatgaggatc 14940
 aagatgcact tttcgcgtat actaagcgta atgtcatccc tactataact caaatgaatc 15000
 ttaagtatgc cattagtgc aagaatagag ctgcaccgt agctgggtgc tctatctgta 15060
 gtactatgac aaatagacag tttcatcaga aattattgaa gtcaatagcc gccactagag 15120
 gagctactgt ggtaattgga acaagcaagt tttacggtgg ctggcataat atgttaaaaa 15180
 ctgtttacag tgatgtagaa actccacacc ttatgggttg ggattatcca aaatgtgaca 15240
 gagccatgcc taacatgctt aggataatgg cctctcttgt tcttgctcgc aaacataaca 15300
 cttgctgtaa cttatcacac cgtttctaca ggtagctaa cgagtgtgcg caagtattaa 15360
 gtgagatggg catgtgtggc ggctcactat atgttaaacc aggtggaaca tcatccgggtg 15420
 atgctacaac tgcttatgct aatagtgtct ttaacatttg tcaagctgtt acagccaatg 15480
 taaatgcact tctttcaact gatggtaata agatagctga caagtatgtc cgcaatctac 15540
 aacacaggct ctatgagtgt ctctatagaa atagggatgt tgatcatgaa ttcgtggatg 15600
 agttttacgc ttacctgctg aaacatttct ccatgatgat tctttctgat gatgccgttg 15660
 tgtgctataa cagtaactat gcggctcaag gtttagtagc tagcattaag aactttaagg 15720
 cagttcttta ttatcaaaat aatgtgttca tgtctgaggc aaaatgttgg actgagactg 15780
 accttactaa aggacctcac gaattttgct cacagcatac aatgctagtt aaacaaggag 15840
 atgattacgt gtacctgcct taccagatc catcaagaat attaggcgca ggctgttttg 15900

S226CAS108.ST25

tcgatgatat tgtcaaaaca gatggtacac ttatgattga aagggttcgtg tcactggcta 15960
ttgatgctta cccacttaca aaacatccta atcaggagta tgctgatgtc tttcacttgt 16020
atttacaata cattagaaag ttacatgatg agcttactgg ccacatgttg gacatgtatt 16080
ccgtaatgct aactaatgat aacacctcac ggtactggga acctgagttt tatgaggcta 16140
tgtacacacc acatacagtc ttgcaggctg taggtgcttg tgtattgtgc aattcacaga 16200
cttcacttcg ttgcggtgcc tgtattagga gaccattcct atgttgcaag tgctgctatg 16260
accatgtcat ttcaacatca cacaattag tggtgtctgt taatccctat gtttgcaatg 16320
ccccagggtg tgatgtcact gatgtgacac aactgtatct aggaggtatg agctattatt 16380
gcaagtcaca taagcctccc attagttttc cattatgtgc taatggtcag gtttttggtt 16440
tatacaaaaa cacatgtgta ggcagtgaca atgtcactga cttcaatgag atagcaacat 16500
gtgattggac taatgctggc gattacatac ttgccaacac ttgtactgag agactcaagc 16560
ttttcgcagc agaaacgctc aaagccactg aggaaacatt taagctgtca tatggtattg 16620
ccactgtacg cgaagtactc tctgacagag aattgcatct ttcattggag gttggaaaac 16680
ctagaccacc attgaacaga aactatgtct ttactgggta ccgtgtaact aaaaatagta 16740
aagtacagat tggagagtac acctttgaaa aagggtgacta tgggtgatgct gttgtgtaca 16800
gaggtactac gacatacaag ttgaatgttg gtgattactt tgtgttgaca tctcacactg 16860
taatgccact tagtgcacct actctagtgc cacaagagca ctatgtgaga attactggct 16920
tgtaccaac actcaacatc tcagatgagt tttctagcaa tgttgcaaat tatcaaaagg 16980
tcggcatgca aaagtactct acactccaag gaccacctgg tactggtaag agtcattttg 17040
ccatcggaact tgctctctat taccatctg ctgcatagt gtatacgga tgctctcatg 17100
cagctgttga tgccctatgt gaaaaggcat taaaatattt gcccatagat aaatgtagta 17160
gaatcatacc tgcgcgtgag cgcgtagagt gttttgataa attcaaagtg aattcaacac 17220
tagaacagta tgttttctgc actgtaaatg cattgccaga aacaactgct gacattgtag 17280
tctttgatga aatctctatg gctactaatt atgacttgag tgttgtcaat gctagacttc 17340
gtgcaaaaca ctacgtctat attggcgatc ctgctcaatt accagcccc cgacattgc 17400
tgactaaagg cacactagaa ccagaatatt ttaattcagt gtgcagactt atgaaaacaa 17460
taggtccaga catgttcctt ggaacttgct gccgttgctc tgctgaaatt gttgacactg 17520
tgagtgcttt agtttatgac aataagctaa aagcacacaa ggataagtca gctcaatgct 17580
tcaaaatgtt ctacaaagggt gttattacac atgatgtttc atctgcaatc aacagacctc 17640
aaataggcgt tgtaagagaa tttcttacac gcaatcctgc ttggagaaaa gctgttttta 17700
tctcacctta taattcacag aacgctgtag cttcaaaaat cttaggattg cctacgcaga 17760
ctgttgattc atcacagggt tctgaatatg actatgtcat attcacacaa actactgaaa 17820
cagcacactc ttgtaatgtc aaccgcttca atgtggctat cacaaggga aaaattggca 17880
ttttgtgcat aatgtctgat agagatcttt atgacaaact gcaatttaca agtctagaaa 17940

S226CAS108.ST25

taccacgtcg	caatgtggct	acattacaag	cagaaaatgt	aactggactt	tttaaggact	18000
gtagtaagat	cattactggg	cttcacaccta	cacaggcacc	tacacacctc	agcgttgata	18060
taaagttcaa	gactgaagga	ttatgtgttg	acataccagg	cataccaaag	gacatgacct	18120
accgtagact	catctctatg	atgggtttca	aaatgaatta	ccaagtcaat	ggttacccta	18180
atatgtttat	cacccgcgaa	gaagctattc	gtcacgttcg	tgcgtggatt	ggctttgatg	18240
tagagggctg	tcagtcaact	agagatgctg	tgggtactaa	cctacctctc	cagctaggat	18300
tttctacagg	tgtaactta	gtagctgtac	cgactgggta	tggtgacact	gaaaataaca	18360
cagaattcac	cagagttaat	gcaaacctc	caccagggtga	ccagttttaa	catcttatac	18420
cactcatgta	taaaggcttg	ccctggaatg	tagtgcgtat	taagatagta	caaatgctca	18480
gtgatacact	gaaaggattg	tcagacagag	tcgtgttcgt	cctttgggcg	catggctttg	18540
agcttacatc	aatgaagtac	tttgtcaaga	ttggacctga	aagaacgtgt	tgtctgtgtg	18600
acaaacgtgc	aacttgcttt	tctacttcat	cagatactta	tgcctgctgg	aatcattctg	18660
tgggttttga	ctatgtctat	aaccatttta	tgattgatgt	tcagcagtgg	ggctttacgg	18720
gtaaccttca	gagtaaccat	gaccaacatt	gccagggtaca	tggaaatgca	catgtggcta	18780
gttgtgatgc	tatcatgact	agatgtttag	cagtccatga	gtgctttgtt	aagcgcggtg	18840
attggtctgt	tgaataccct	attataggag	atgaactgag	ggttaattct	gcttgacaaa	18900
aagtacaaca	catgggttgtg	aagtctgcat	tgcttgctga	taagtttcca	gttcttcatg	18960
acattggaaa	tcaaaggct	atcaagtgtg	tgccctaggc	tgaagtagaa	tggaagtctt	19020
acgatgctca	gccatgtagt	gacaaagctt	acaaaataga	ggaactcttc	tattcttatg	19080
ctacacatca	cgataaattc	actgatgggtg	tttgtttgtt	ttggaattgt	aacgttgatc	19140
gttaccagc	caatgcaatt	gtgtgtaggt	ttgacacaag	agtcttgtca	aacttgaact	19200
taccaggctg	tgatgggtgg	agtttgtatg	tgaataagca	tgcattccac	actccagctt	19260
tcgataaaag	tgcatcttact	aatttaaagc	aattgccttt	cttttactat	tctgatagtc	19320
cttgtgagtc	tcatggcaaa	caagtagtgt	cggatattga	ttatgttcca	ctcaaactctg	19380
ctacgtgtat	tacacgatgc	aatttaggtg	gtgctgtttg	cagacaccat	gcaaatgagt	19440
accgacagta	cttggaatgca	tataatatga	tgatttctgc	tggatttagc	ctatggattt	19500
acaaacaatt	tgatacttat	aacctgtgga	atacatttac	caggttacag	agtttagaaa	19560
atgtggctta	taatgttgtt	aataaaggac	actttgatgg	acacgccggc	gaagcacctg	19620
tttccatcat	taataatgct	gtttacacaa	aggtagatgg	tattgatgtg	gagatctttg	19680
aaaataagac	aacacttcct	gttaatgttg	catttgagct	ttgggctaag	cgtaacatta	19740
aaccagtgcc	agagattaag	atactcaata	atttgggtgt	tgatatcgct	gctaatactg	19800
taatctggga	ctacaaaaga	gaagccccag	cacatgtatc	tacaatagg	gtctgcacaa	19860
tgactgacat	tgccaagaaa	cctactgaga	gtgctgttcc	ttcacttact	gtcttgtttg	19920
atggtagagt	ggaaggacag	gtagaccttt	ttagaaacgc	ccgtaatggg	gttttaataa	19980

S226CAS108.ST25

cagaaggttc	agtcaaaggt	ctaacacctt	caaagggacc	agcacaagct	agcgtcaatg	20040
gagtcacatt	aattggagaa	tcagtaaaaa	cacagtttaa	ctactttaag	aaagtagacg	20100
gcattattca	acagttgcct	gaaacctact	ttactcagag	cagagactta	gaggatttta	20160
agcccagatc	acaaatggaa	actgactttc	tcgagctcgc	tatggatgaa	ttcatacagc	20220
gatataagct	cgagggctat	gccttcgaac	acatcgttta	tggagatttc	agtcattggac	20280
aacttggcgg	tcttcattta	atgataggct	tagccaagcg	ctcacaagat	tcaccactta	20340
aattagagga	ttttatccct	atggacagca	cagtgaaaaa	ttacttcata	acagatgcgc	20400
aaacaggttc	atcaaaatgt	gtgtgttctg	tgattgatct	tttacttgat	gactttgtcg	20460
agataataaa	gtcacaagat	ttgtcagtga	tttcaaaagt	ggtcaagggt	acaattgact	20520
atgctgaaat	ttcattcatg	ctttggtgta	aggatggaca	tggtgaaacc	ttctacccaa	20580
aactacaagc	aagtcaagcg	tggcaaccag	gtgttgcatg	gcctaacttg	tacaagatgc	20640
aaagaatgct	tcttgaaaag	tgtgaccttc	agaattatgg	tgaaaatgct	gttataccaa	20700
aaggaataat	gatgaatgtc	gcaaagtata	ctcaactgtg	tcaatactta	aatacactta	20760
ctttagctgt	accctacaac	atgagagtta	ttcactttgg	tgctggctct	gataaaggag	20820
ttgcaccagg	tacagctgtg	ctcagacaat	ggttgccaac	tggcacacta	cttgctgatt	20880
cagatcttaa	tgacttcgtc	tccgacgcag	attctacttt	aattggagac	tgtgcaacag	20940
tacatacggc	taataaatgg	gaccttatta	ttagcgatat	gtatgaccct	aggaccaaac	21000
atgtgacaaa	agagaatgac	tctaaagaag	ggtttttcac	ttatctgtgt	ggatttataa	21060
agcaaaaact	agccctgggt	ggttctatag	ctgtaaagat	aacagagcat	tcttggaatg	21120
ctgaccttta	caagcttatg	ggccatttct	catggtggac	agcttttggt	acaaatgtaa	21180
atgcatcatc	atcggaagca	tttttaattg	gggctaacta	tcttggcaag	ccgaaggaa	21240
aaattgatgg	ctataccatg	catgctaact	acattttctg	gaggaacaca	aatcctatcc	21300
agttgtcttc	ctattcactc	tttgacatga	gcaaatttcc	tcttaaatta	agaggaactg	21360
ctgtaatgtc	tcttaaggag	aatcaaatca	atgatatgat	ttattctctt	ctggaaaaag	21420
gtaggcttat	cattagagaa	aacaacagag	ttgtgggtttc	aagtgatatt	cttggttaaca	21480
actaaacgaa	catgtttatt	ttcttattat	ttcttactct	cactagtggg	agtgaccttg	21540
accggtgcac	cacttttgat	gatgttcaag	ctcctaatta	cactcaacat	acttcatcta	21600
tgaggggggt	ttactatcct	gatgaaattt	ttagatcaga	cactctttat	ttactcagg	21660
atttatttct	tccattttat	tctaattgta	cagggtttca	tactattaat	catacgtttg	21720
gcaaccctgt	catacctttt	aaggatggta	tttattttgc	tgccacagag	aaatcaaagt	21780
ttgtccgtgg	ttgggttttt	ggttctacca	tgaacaacaa	gtcacagtcg	gtgattatta	21840
ttaacaattc	tactaatgtt	gttatacgag	catgtaactt	tgaattgtgt	gacaaccctt	21900
tctttgctgt	ttctaaaccc	atgggtacac	agacacatac	tatgatattc	gataatgcat	21960
ttaattgcac	tttcgagtac	atatctgatg	ccttttcgct	tgatgtttca	gaaaagtcag	22020

S226CAS108.ST25

gtaatttttaa acacttacga gagtttgtgt ttaaaaaataa agatggggttt ctctatgttt 22080
 ataagggcta tcaacctata gatgtagttc gtgatctacc ttctgggtttt aacactttga 22140
 aacctatttt taagttgcct cttggtatta acattacaaa ttttagagcc attcttacag 22200
 ccttttcacc tgctcaagac atttggggca cgtcagctgc agcctatttt gttggctatt 22260
 taaagccaac tacatttatg ctcaagtatg atgaaaatgg tacaatcaca gatgctgttg 22320
 attgttctca aaatccactt gctgaactca aatgctctgt taagagcttt gagattgaca 22380
 aaggaattta ccagacctct aatttcaggg ttgttccctc aggagatgtt gtgagattcc 22440
 ctaatattac aaacttgtgt ccttttggag aggtttttta tgctactaaa ttcccttctg 22500
 tctatgcatg ggagagaaaa aaaatttcta attgtgttgc tgattactct gtgctctaca 22560
 actcaacatt tttttcaacc ttttaagtgc atggcgtttc tgccactaag ttgaatgatc 22620
 tttgcttctc caatgtctat gcagattctt ttgtagtcaa gggagatgat gtaagacaaa 22680
 tagcgccagg acaaactggg gttattgctg attataatta taaattgcca gatgatttca 22740
 tgggttgtgt ccttgcttgg aatactagga acattgatgc tacttcaact ggtaattata 22800
 attataaata taggtatctt agacatggca agcttaggcc ctttgagaga gacatatcta 22860
 atgtgccttt ctcccctgat ggcaaacctt gcacccacc tgctcttaat tgttattggc 22920
 cattaaatga ttatgggttt tacaccacta ctggcattgg ctaccaacct tacagagttg 22980
 tagtactttc ttttgaactt ttaaattgcac cggccacggg ttgtggacca aaattatcca 23040
 ctgaccttat taagaaccag tgtgtcaatt ttaattttta tggactcact ggtactggtg 23100
 tgttaactcc ttcttcaaag agatttcaac catttcaaca atttggccgt gatgtttctg 23160
 atttactga ttccgttcga gatcctaaaa catctgaaat attagacatt tcaccttgct 23220
 cttttggggg tgtaagtgtg attacacctg gaacaaatgc ttcacttgaa gttgctgttc 23280
 tatatcaaga tgttaactgc actgatgttt ctacagcaat tcatgcagat caactcacac 23340
 cagcttggcg catatattct actggaaaca atgtattcca gactcaagca ggctgtctta 23400
 taggagctga gcatgtcgac acttcttatg agtgcgacat tcctattgga gctggcattt 23460
 gtgctagtta ccatacagtt tctttattac gtagtactag ccaaaaatct attgtggctt 23520
 atactatgtc tttaggtgct gatagttcaa ttgcttactc taataacacc attgctatac 23580
 ctactaactt ttcaattagc attactacag aagtaatgcc tgtttctatg gctaaaacct 23640
 ccgtagattg taatatgtac atctgcggag atttctactga atgtgctaatt ttgcttctcc 23700
 aatatggtag cttttgcaca caactaaatc gtgcactctc aggtattgct gctgaacagg 23760
 atcgcaacac acgtgaagtg ttcgctcaag tcaaacaaat gtacaaaacc ccaactttga 23820
 aatatttttg tggttttaat ttttcacaaa tattacctga ccttctaaag ccaactaaga 23880
 ggtcttttat tgaggacttg ctctttaata aggtgacact cgctgatgct ggcttcatga 23940
 agcaatatgg cgaatgccta ggtgatatta atgctagaga tctcatttgt gcgcagaagt 24000
 tcaatggact tacagtgttg ccacctctgc tcaatgatga tatgattgct gcctacactg 24060

S226CAS108.ST25

ctgctctagt tagtggtact gccactgctg gatggacatt tgggtgctggc gctgctcttc 24120
 aaataccttt tgctatgcaa atggcatata ggttcaatgg cattggagtt acccaaaatg 24180
 ttctctatga gaaccaaaaa caaatcgcca accaatttaa caaggcgatt agtcaaattc 24240
 aagaatcact tacaacaaca tcaactgcat tgggcaagct gcaagacgtt gttaaccaga 24300
 atgctcaagc attaaacaca cttgttaaac aacttagctc taattttggt gcaatttcaa 24360
 gtgtgctaaa tgatatcctt tcgcgacttg ataaagtcga ggcggaggta caaattgaca 24420
 ggtaattac aggcagactt caaagccttc aaacctatgt aacacaacaa ctaatcaggg 24480
 ctgctgaaat cagggcttct gctaactctg ctgctactaa aatgtctgag tgtgttcttg 24540
 gacaatcaaa aagagttgac ttttgtggaa agggctacca cttatgtcc ttcccacaag 24600
 cagccccgca tgggtgttgc ttctacatg tcacgtatgt gccatcccag gagaggaact 24660
 tcaccacagc gccagcaatt tgtcatgaag gcaaagcata cttccctcgt gaagggtgtt 24720
 ttgtgtttaa tggcacttct tggtttatta cacagaggaa cttcttttct ccacaaataa 24780
 ttactacaga caatacattt gtctcaggaa atttgtatgt cgttattggc atcattaaca 24840
 acacagttaa tgatcctctg caacctgagc ttgactcatt caaagaagag ctggacaagt 24900
 acttcaaaaa tcatacatca ccagatgttg atcttggcga catttcaggc attaacgctt 24960
 ctgtcgtcaa cattcaaaaa gaaattgacc gcctcaatga ggtcgctaaa aatttaaagt 25020
 aatcactcat tgaccttcaa gaattgggaa aatatgagca atatattaaa tggccttggt 25080
 atgtttggct cggcttcatt gctggactaa ttgccatcgt catggttaca atcttgcttt 25140
 gttgcatgac tagttgttgc agttgcctca aggggtgatg ctcttggtgt tcttgctgca 25200
 agtttgatga ggatgactct gagccagttc tcaagggtgt caaattacat tacacataaa 25260
 cgaacttatg gatttgttta tgagattttt tactcttggg tcaattactg cacagccagt 25320
 aaaaattgac aatgcttctc ctgcaagtac tgttcattgt acagcaacga taccgctaca 25380
 agcctcactc ctttccggat ggcttgttat tggcgttgca tttcttgctg ttttccagag 25440
 cgctacaaa ataattgcgc tcaataaaag atggcagcta gccctttata agggcttcca 25500
 gttcatttgc aatttactgc tgctatttgt taccatctat tcacatcttt tgcttgctgc 25560
 tgcaggatat gaggcgcaat tttgtacct ctatgccttg atatattttc tacaatgcat 25620
 caacgcatgt agaattatta tgagatgttg gctttgttgg aagtgcaaat ccaagaaccc 25680
 attactttat gatgccaaact actttgtttg ctggcacaca cataactatg actactgtat 25740
 accatataac agtgtcacag atacaattgt cgttactgaa ggtgacggca tttcaacacc 25800
 aaaactcaaa gaagactacc aaattggtgg ttattctgag gataggcact caggtgttaa 25860
 agactatgtc gttgtacatg gctatttcac cgaagtttac taccagcttg agtctacaca 25920
 aattactaca gacactggta ttgaaaatgc tacattcttc atctttaaca agcttggttaa 25980
 agaccaccg aatgtgcaaa tacacacaat cgacggctct tcaggagttg ctaatccagc 26040
 aatggatcca atttatgatg agccgacgac gactactagc gtgcctttgt aagcacaaga 26100

S226CAS108.ST25

aagtgagtac gaacttatgt actcattcgt ttcggaagaa acagggtacgt taatagttaa 26160
tagcgtagctt ctttttcttg ctttcgtggg attcttgcta gtcacactag ccatccttac 26220
tgcgcttcga ttgtgtgcgt actgctgcaa tattgttaac gtgagtttag taaaaccaac 26280
ggtttacgtc tactcgctg ttaaaaatct gaactcttct gaaggagttc ctgatcttct 26340
ggctctaaacg aactaactat tattattatt ctgtttggaa ctttaacatt gcttatcatg 26400
gcagacaacg gtactattac cgttgaggag cttaacaac tcctggaaca atggaaccta 26460
gtaatagggt tcctattcct agcctggatt atgttactac aatttgcccta ttctaactcg 26520
aacagggtttt tgtacataat aaagcttggt ttcctctggc tcttggtggc agtaacactt 26580
gcttgttttg tgcttgctgc tgtctacaga attaattggg tgactggcg gattgcgatt 26640
gcaatggctt gtattgtagg ctgatgtgg cttagctact tcgttgcttc cttcaggctg 26700
tttgctcgtc ccgctcaat gtggctattc aaccagaaa caaacattct tctcaatgtg 26760
cctctccggg ggacaattgt gaccagaccg ctcatggaaa gtgaacttgt cattgggtgct 26820
gtgatcattc gtggtcactt gcgaatggcc ggacactccc tagggcgctg tgacattaag 26880
gacctgcaa aagagatcac tgtggctaca tcacgaacgc tttcttatta caaattagga 26940
gcgtcgcagc gtgtaggcac tgattcaggt tttgctgcat acaaccgcta ccgtattgga 27000
aactataaat taaatacaga ccacgccggt agcaacgaca atattgcttt gctagtacag 27060
taagtgacaa cagatgtttc atcttgttga cttccagggt acaatagcag agatattgat 27120
tatcattatg aggactttca ggattgctat ttggaatctt gacgttataa taagttcaat 27180
agtgagacaa ttatttaagc ctctaactaa gaagaattat tcggagttag atgatgaaga 27240
acctatggag ttagattatc cataaaacga acatgaaaat tattctcttc ctgacattga 27300
ttgtatttac atcttgcgag ctatatcact atcaggagtg tgttagaggt acgactgtac 27360
tactaaaaga accttgccca tcaggaacat acgagggcaa ttcaccattt caccctcttg 27420
ctgacaataa atttgcaacta acttgcaacta gcacacactt tgcttttgct tgtgctgacg 27480
gtactcgaca tacctatcag ctgcgtgcaa gatcagtttc accaaaactt ttcacagac 27540
aagaggaggt tcaacaagag ctctactcgc cacttttctt cattgttgct gctctagtat 27600
ttttaatact ttgcttcacc attaagagaa agacagaatg aatgagctca ctttaattga 27660
cttctatttg tgcttttttag ctttctgct attccttggt ttaataatgc ttattatatt 27720
ttggttttca ctcgaaatcc aggatctaga agaacctgt accaaagtct aaacgaacat 27780
gaaacttctc attgttttga cttgtatttc tctatgcagt tgcatatgca ctgtagtaca 27840
gcgctgtgca tctaataaac ctcatgtgct tgaagatcct tgtaaggtag aacactaggg 27900
gtaatactta tagcactgct tggctttgtg ctctaggaaa ggttttacct tttcatagat 27960
ggcactat ggttcaaaca tgcacaccta atgttactat caactgtcaa gatccagctg 28020
gtggtgctgct tatagctagg tgttggtacc ttcatgaagg tcaccaaact gctgcattta 28080
gagacgtact tgttgtttta aataaacgaa caaattaaaa tgtctgataa tggaccccaa 28140

S226CAS108.ST25

```

tcaaaccaac gtagtgcccc ccgcattaca tttggtggac ccacagattc aactgacaat 28200
aaccagaatg gaggacgcaa tggggcaagg ccaaaacagc gccgaccca aggtttaccc 28260
aataaactg cgtcttggtt cacagctctc actcagcatg gcaaggagga acttagattc 28320
cctcgaggcc agggcggttc aatcaacacc aatagtggtc cagatgacca aattggctac 28380
taccgaagag ctacccgacg agttcgtggt ggtgacggca aaatgaaaga gctcagcccc 28440
agatggtact tctattacct aggaactggc ccagaagctt cacttccta cggcgctaac 28500
aaagaaggca tcgtatgggt tgcaactgag ggagccttga atacaccaa agaccacatt 28560
ggcaccgca atcctaataa caatgctgcc accgtgctac aacttcctca aggaacaaca 28620
ttgccaaaag gcttctacgc agagggaaag agaggcggca gtcaagcctc ttctcgctcc 28680
tcatcacgta gtcgcggtaa ttcaagaaat tcaactcctg gcagcagtag gggaaattct 28740
cctgctcgaa tggctagcgg aggtggtgaa actgccctcg cgctattgct gctagacaga 28800
ttgaaccagc ttgagagcaa agtttctggt aaaggccaac aacaacaagg ccaaactgtc 28860
actaagaaat ctgctgctga ggcattctaa aagcctcgcc aaaaacgtac tgccacaaaa 28920
cagtacaacg tcaactaagc atttgggaga cgtggtccag acaaaacca aggaaatttc 28980
ggggaccaag acctaatac acaaggaact gattacaaac attggccgca aattgcacaa 29040
tttgcctcaa gtgcctctgc attctttgga atgtcacgca ttggcatgga agtcacacct 29100
tcgggaacat ggctgactta tcatggagcc attaaattgg atgacaaaga tccacaattc 29160
aaagacaacg tcatactgct gaacaagcac attgacgcat acaaaacatt cccaccaaca 29220
gagcctaaaa aggacaaaaa gaaaaagact gatgaagctc agcctttgcc gcagagacaa 29280
aagaagcagc ccaactgtgac tcttcttcct gcggctgaca tggatgattt ctccagacaa 29340
cttcaaaatt ccatgagtgg agcttctgct gattcaactc aggcataaac actcatgatg 29400
accacacaag gcagatgggc tatgtaaacg ttttcgcaat tccgtttacg atacatagtc 29460
tactcttggt cagaatgaat tctcgttaact aaacagcaca agtaggttta gttaacttta 29520
atctcacata gcaatcttta atcaatgtgt aacattaggg aggacttgaa agagccacca 29580
cattttcatc gaggccacgc ggagtacgat cgagggtaca gtgaataatg ctaggagagag 29640
ctgcctatat ggaagagccc taatgtgtaa aattaatttt agtagtgcta tccccatgtg 29700
attttaatag cttcttagga gaatgacaaa aaaaaaaaaa aaaaaa 29746

```

<210> 2

<211> 3945

<212> DNA

<213> CORONAVIRUS

<220>

S226CAS108.ST25

<221> CDS

<222> (89)..(3853)

<223>

<400> 2

ttctcttctg gaaaaaggta ggcttatcat tagagaaaac aacagagttg tggtttcaag 60

tgatattctt	gtaacaact	aaacgaac	atg	ttt	att	ttc	tta	tta	ttt	ctt	112
			Met	Phe	Ile	Phe	Leu	Leu	Phe	Leu	
			1				5				

act	ctc	act	agt	ggg	agt	gac	ctt	gac	cgg	tgc	acc	act	ttt	gat	gat	160
Thr	Leu	Thr	Ser	Gly	Ser	Asp	Leu	Asp	Arg	Cys	Thr	Thr	Phe	Asp	Asp	
	10					15					20					

gtt	caa	gct	cct	aat	tac	act	caa	cat	act	tca	tct	atg	agg	ggg	gtt	208
Val	Gln	Ala	Pro	Asn	Tyr	Thr	Gln	His	Thr	Ser	Ser	Met	Arg	Gly	Val	
25					30					35					40	

tac	tat	cct	gat	gaa	att	ttt	aga	tca	gac	act	ctt	tat	tta	act	cag	256
Tyr	Tyr	Pro	Asp	Glu	Ile	Phe	Arg	Ser	Asp	Thr	Leu	Tyr	Leu	Thr	Gln	
				45					50					55		

gat	tta	ttt	ctt	cca	ttt	tat	tct	aat	gtt	aca	ggg	ttt	cat	act	att	304
Asp	Leu	Phe	Leu	Pro	Phe	Tyr	Ser	Asn	Val	Thr	Gly	Phe	His	Thr	Ile	
			60					65					70			

aat	cat	acg	ttt	ggc	aac	cct	gtc	ata	cct	ttt	aag	gat	ggg	att	tat	352
Asn	His	Thr	Phe	Gly	Asn	Pro	Val	Ile	Pro	Phe	Lys	Asp	Gly	Ile	Tyr	
		75					80				85					

ttt	gct	gcc	aca	gag	aaa	tca	aat	gtt	gtc	cgt	ggg	tgg	gtt	ttt	ggg	400
Phe	Ala	Ala	Thr	Glu	Lys	Ser	Asn	Val	Val	Arg	Gly	Trp	Val	Phe	Gly	
	90					95					100					

tct	acc	atg	aac	aac	aag	tca	cag	tcg	gtg	att	att	att	aac	aat	tct	448
Ser	Thr	Met	Asn	Asn	Lys	Ser	Gln	Ser	Val	Ile	Ile	Ile	Asn	Asn	Ser	
105					110					115					120	

act	aat	gtt	gtt	ata	cga	gca	tgt	aac	ttt	gaa	ttg	tgt	gac	aac	cct	496
Thr	Asn	Val	Val	Ile	Arg	Ala	Cys	Asn	Phe	Glu	Leu	Cys	Asp	Asn	Pro	
				125					130					135		

ttc	ttt	gct	gtt	tct	aaa	ccc	atg	ggg	aca	cag	aca	cat	act	atg	ata	544
Phe	Phe	Ala	Val	Ser	Lys	Pro	Met	Gly	Thr	Gln	Thr	His	Thr	Met	Ile	
			140					145					150			

ttc	gat	aat	gca	ttt	aat	tgc	act	ttc	gag	tac	ata	tct	gat	gcc	ttt	592
Phe	Asp	Asn	Ala	Phe	Asn	Cys	Thr	Phe	Glu	Tyr	Ile	Ser	Asp	Ala	Phe	
		155					160					165				

tcg	ctt	gat	gtt	tca	gaa	aag	tca	ggg	aat	ttt	aaa	cac	tta	cga	gag	640
Ser	Leu	Asp	Val	Ser	Glu	Lys	Ser	Gly	Asn	Phe	Lys	His	Leu	Arg	Glu	
	170					175					180					

ttt	gtg	ttt	aaa	aat	aaa	gat	ggg	ttt	ctc	tat	gtt	tat	aag	ggc	tat	688
Phe	Val	Phe	Lys	Asn	Lys	Asp	Gly	Phe	Leu	Tyr	Val	Tyr	Lys	Gly	Tyr	
185					190				195						200	

caa	cct	ata	gat	gta	gtt	cgt	gat	cta	cct	tct	ggg	ttt	aac	act	ttg	736
Gln	Pro	Ile	Asp	Val	Val	Arg	Asp	Leu	Pro	Ser	Gly	Phe	Asn	Thr	Leu	
				205					210					215		

aaa	cct	att	ttt	aag	ttg	cct	ctt	ggg	att	aac	att	aca	aat	ttt	aga	784
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

S226CAS108.ST25															
Lys	Pro	Ile	Phe	Lys	Leu	Pro	Leu	Gly	Ile	Asn	Ile	Thr	Asn	Phe	Arg
			220					225					230		
gcc	att	ctt	aca	gcc	ttt	tca	cct	gct	caa	gac	att	tgg	ggc	acg	tca
Ala	Ile	Leu	Thr	Ala	Phe	Ser	Pro	Ala	Gln	Asp	Ile	Trp	Gly	Thr	Ser
		235					240					245			
gct	gca	gcc	tat	ttt	gtt	ggc	tat	tta	aag	cca	act	aca	ttt	atg	ctc
Ala	Ala	Ala	Tyr	Phe	Val	Gly	Tyr	Leu	Lys	Pro	Thr	Thr	Phe	Met	Leu
		250				255					260				
aag	tat	gat	gaa	aat	ggc	aca	atc	aca	gat	gct	gtt	gat	tgt	tct	caa
Lys	Tyr	Asp	Glu	Asn	Gly	Thr	Ile	Thr	Asp	Ala	Val	Asp	Cys	Ser	Gln
265					270					275					280
aat	cca	ctt	gct	gaa	ctc	aaa	tgc	tct	gtt	aag	agc	ttt	gag	att	gac
Asn	Pro	Leu	Ala	Glu	Leu	Lys	Cys	Ser	Val	Lys	Ser	Phe	Glu	Ile	Asp
				285					290					295	
aaa	gga	att	tac	cag	acc	tct	aat	ttc	agg	gtt	gtt	ccc	tca	gga	gat
Lys	Gly	Ile	Tyr	Gln	Thr	Ser	Asn	Phe	Arg	Val	Val	Pro	Ser	Gly	Asp
			300					305					310		
gtt	gtg	aga	ttc	cct	aat	att	aca	aac	ttg	tgt	cct	ttt	gga	gag	gtt
Val	Val	Arg	Phe	Pro	Asn	Ile	Thr	Asn	Leu	Cys	Pro	Phe	Gly	Glu	Val
		315					320					325			
ttt	aat	gct	act	aaa	ttc	cct	tct	gtc	tat	gca	tgg	gag	aga	aaa	aaa
Phe	Asn	Ala	Thr	Lys	Phe	Pro	Ser	Val	Tyr	Ala	Trp	Glu	Arg	Lys	Lys
	330					335					340				
att	tct	aat	tgt	gtt	gct	gat	tac	tct	gtg	ctc	tac	aac	tca	aca	ttt
Ile	Ser	Asn	Cys	Val	Ala	Asp	Tyr	Ser	Val	Leu	Tyr	Asn	Ser	Thr	Phe
345					350					355					360
ttt	tca	acc	ttt	aag	tgc	tat	ggc	gtt	tct	gcc	act	aag	ttg	aat	gat
Phe	Ser	Thr	Phe	Lys	Cys	Tyr	Gly	Val	Ser	Ala	Thr	Lys	Leu	Asn	Asp
				365					370					375	
ctt	tgc	ttc	tcc	aat	gtc	tat	gca	gat	tct	ttt	gta	gtc	aag	gga	gat
Leu	Cys	Phe	Ser	Asn	Val	Tyr	Ala	Asp	Ser	Phe	Val	Val	Lys	Gly	Asp
			380					385					390		
gat	gta	aga	caa	ata	gcg	cca	gga	caa	act	ggc	gtt	att	gct	gat	tat
Asp	Val	Arg	Gln	Ile	Ala	Pro	Gly	Gln	Thr	Gly	Val	Ile	Ala	Asp	Tyr
		395					400					405			
aat	tat	aaa	ttg	cca	gat	gat	ttc	atg	ggc	tgt	gtc	ctt	gct	ttg	aat
Asn	Tyr	Lys	Leu	Pro	Asp	Asp	Phe	Met	Gly	Cys	Val	Leu	Ala	Trp	Asn
	410					415					420				
act	agg	aac	att	gat	gct	act	tca	act	ggc	aat	tat	aat	tat	aaa	tat
Thr	Arg	Asn	Ile	Asp	Ala	Thr	Ser	Thr	Gly	Asn	Tyr	Asn	Tyr	Lys	Tyr
425					430					435					440
agg	tat	ctt	aga	cat	ggc	aag	ctt	agg	ccc	ttt	gag	aga	gac	ata	tct
Arg	Tyr	Leu	Arg	His	Gly	Lys	Leu	Arg	Pro	Phe	Glu	Arg	Asp	Ile	Ser
				445					450					455	
aat	gtg	cct	ttc	tcc	cct	gat	ggc	aaa	cct	tgc	acc	cca	cct	gct	ctt
Asn	Val	Pro	Phe	Ser	Pro	Asp	Gly	Lys	Pro	Cys	Thr	Pro	Pro	Ala	Leu
			460					465					470		
aat	tgt	tat	tgg	cca	tta	aat	gat	tat	ggc	ttt	tac	acc	act	act	ggc
Asn	Cys	Tyr	Trp	Pro	Leu	Asn	Asp	Tyr	Gly	Phe	Tyr	Thr	Thr	Thr	Gly
		475					480					485			
att	ggc	tac	caa	cct	tac	aga	gtt	gta	gta	ctt	tct	ttt	gaa	ctt	tta

ile	Gly 490	Tyr	Gln	Pro	Tyr	Arg 495	Val	Val	Val	Leu	Ser 500	Phe	Glu	Leu	Leu	
aat	gca	ccg	gcc	acg	gtt	tgt	gga	cca	aaa	tta	tcc	act	gac	ctt	att	1648
Asn	Ala	Pro	Ala	Thr	Val	Cys	Gly	Pro	Lys	Leu	Ser	Thr	Asp	Leu	Ile	520
505					510					515						
aag	aac	cag	tgt	gtc	aat	ttt	aat	ttt	aat	gga	ctc	act	ggt	act	ggt	1696
Lys	Asn	Gln	Cys	Val	Asn	Phe	Asn	Phe	Asn	Gly	Leu	Thr	Gly	Thr	Gly	535
				525					530							
gtg	tta	act	cct	tct	tca	aag	aga	ttt	caa	cca	ttt	caa	caa	ttt	ggc	1744
Val	Leu	Thr	Pro	Ser	Ser	Lys	Arg	Phe	Gln	Pro	Phe	Gln	Gln	Phe	Gly	550
			540					545								
cgt	gat	gtt	tct	gat	ttc	act	gat	tcc	gtt	cga	gat	cct	aaa	aca	tct	1792
Arg	Asp	Val	Ser	Asp	Phe	Thr	Asp	Ser	Val	Arg	Asp	Pro	Lys	Thr	Ser	565
		555					560									
gaa	ata	tta	gac	att	tca	cct	tgc	tct	ttt	ggg	ggt	gta	agt	gta	att	1840
Glu	Ile	Leu	Asp	Ile	Ser	Pro	Cys	Ser	Phe	Gly	Gly	Val	Ser	Val	Ile	580
	570					575										
aca	cct	gga	aca	aat	gct	tca	tct	gaa	gtt	gct	gtt	cta	tat	caa	gat	1888
Thr	Pro	Gly	Thr	Asn	Ala	Ser	Ser	Glu	Val	Ala	Val	Leu	Tyr	Gln	Asp	600
					590					595						
gtt	aac	tgc	act	gat	gtt	tct	aca	gca	att	cat	gca	gat	caa	ctc	aca	1936
Val	Asn	Cys	Thr	Asp	Val	Ser	Thr	Ala	Ile	His	Ala	Asp	Gln	Leu	Thr	615
				605					610							
cca	gct	tgg	cgc	ata	tat	tct	act	gga	aac	aat	gta	ttc	cag	act	caa	1984
Pro	Ala	Trp	Arg	Ile	Tyr	Ser	Thr	Gly	Asn	Asn	Val	Phe	Gln	Thr	Gln	625
			620					625					630			
gca	ggc	tgt	ctt	ata	gga	gct	gag	cat	gtc	gac	act	tct	tat	gag	tgc	2032
Ala	Gly	Cys	Leu	Ile	Gly	Ala	Glu	His	Val	Asp	Thr	Ser	Tyr	Glu	Cys	640
		635					640					645				
gac	att	cct	att	gga	gct	ggc	att	tgt	gct	agt	tac	cat	aca	gtt	tct	2080
Asp	Ile	Pro	Ile	Gly	Ala	Gly	Ile	Cys	Ala	Ser	Tyr	His	Thr	Val	Ser	650
	650					655					660					
tta	tta	cgt	agt	act	agc	caa	aaa	tct	att	gtg	gct	tat	act	atg	tct	2128
Leu	Leu	Arg	Ser	Thr	Ser	Gln	Lys	Ser	Ile	Val	Ala	Tyr	Thr	Met	Ser	665
					670					675					680	
tta	ggt	gct	gat	agt	tca	att	gct	tac	tct	aat	aac	acc	att	gct	ata	2176
Leu	Gly	Ala	Asp	Ser	Ser	Ile	Ala	Tyr	Ser	Asn	Asn	Thr	Ile	Ala	Ile	690
				685					690					695		
cct	act	aac	ttt	tca	att	agc	att	act	aca	gaa	gta	atg	cct	gtt	tct	2224
Pro	Thr	Asn	Phe	Ser	Ile	Ser	Ile	Thr	Thr	Glu	Val	Met	Pro	Val	Ser	700
			700					705					710			
atg	gct	aaa	acc	tcc	gta	gat	tgt	aat	atg	tac	atc	tgc	gga	gat	tct	2272
Met	Ala	Lys	Thr	Ser	Val	Asp	Cys	Asn	Met	Tyr	Ile	Cys	Gly	Asp	Ser	715
		715					720					725				
act	gaa	tgt	gct	aat	ttg	ctt	ctc	caa	tat	ggt	agc	ttt	tgc	aca	caa	2320
Thr	Glu	Cys	Ala	Asn	Leu	Leu	Leu	Gln	Tyr	Gly	Ser	Phe	Cys	Thr	Gln	730
		730				735					740					
cta	aat	cgt	gca	ctc	tca	ggt	att	gct	gct	gaa	cag	gat	cgc	aac	aca	2368
Leu	Asn	Arg	Ala	Leu	Ser	Gly	Ile	Ala	Ala	Glu	Gln	Asp	Arg	Asn	Thr	745
					750					755					760	
cgt	gaa	gtg	ttc	gct	caa	gtc	aaa	caa	atg	tac	aaa	acc	cca	act	ttg	2416

S226CAS108.ST25															
Arg	Glu	Val	Phe	Ala 765	Gln	Val	Lys	Gln 770	Met	Tyr	Lys	Thr	Pro	Thr 775	Leu
aaa Lys	tat Tyr	ttt Phe	ggt Gly 780	ggt Gly	ttt Phe	aat Asn	ttt Phe	tca Ser 785	caa Gln	ata Ile	tta Leu	cct Pro	gac Asp 790	cct Pro	cta Leu
aag Lys	cca Pro	act Thr 795	aag Lys	agg Arg	tct Ser	ttt Phe	att Ile 800	gag Glu	gac Asp	ttg Leu	ctc Leu	ttt Phe 805	aat Asn	aag Lys	gtg Val
aca Thr 810	ctc Leu	gct Ala	gat Asp	gct Ala	ggc Gly	ttc Phe 815	atg Met	aag Lys	caa Gln	tat Tyr	ggc Gly 820	gaa Glu	tgc Cys	cta Leu	ggt Gly
gat Asp 825	att Ile	aat Asn	gct Ala	aga Arg	gat Asp 830	ctc Leu	att Ile	tgt Cys	gcg Ala	cag Gln 835	aag Lys	ttc Phe	aat Asn	gga Gly	ctt Leu 840
aca Thr	gtg Val	ttg Leu	cca Pro	cct Pro 845	ctg Leu	ctc Leu	act Thr	gat Asp	gat Asp 850	atg Met	att Ile	gct Ala	gcc Ala	tac Tyr 855	act Thr
gct Ala	gct Ala	cta Leu	gtt Val 860	agt Ser	ggc Gly	act Thr	gcc Ala	act Thr 865	gct Ala	gga Gly	tgg Trp	aca Thr	ttt Phe 870	ggt Gly	gct Ala
ggc Gly	gct Ala	gct Ala 875	ctt Leu	caa Gln	ata Ile	cct Pro	ttt Phe 880	gct Ala	atg Met	caa Gln	atg Met	gca Ala 885	tat Tyr	agg Arg	ttc Phe
aat Asn 890	ggc Gly	att Ile	gga Gly	gtt Val	acc Thr	caa Gln 895	aat Asn	gtt Val	ctc Leu	tat Tyr	gag Glu 900	aac Asn	caa Gln	aaa Lys	caa Gln
atc Ile 905	gcc Ala	aac Asn	caa Gln	ttt Phe	aac Asn 910	aag Lys	gcg Ala	att Ile	agt Ser	caa Gln 915	att Ile	caa Gln	gaa Glu	tca Ser	ctt Leu 920
aca Thr	aca Thr	aca Thr	tca Ser	act Thr 925	gca Ala	ttg Leu	ggc Gly	aag Lys	ctg Leu 930	caa Gln	gac Asp	gtt Val	gtt Val	aac Asn 935	cag Gln
aat Asn	gct Ala	caa Gln	gca Ala 940	tta Leu	aac Asn	aca Thr	ctt Leu	gtt Val 945	aaa Lys	caa Gln	ctt Leu	agc Ser	tct Ser 950	aat Asn	ttt Phe
ggt Gly	gca Ala	att Ile 955	tca Ser	agt Ser	gtg Val	cta Leu	aat Asn 960	gat Asp	atc Ile	ctt Leu	tcg Ser	cga Arg 965	ctt Leu	gat Asp	aaa Lys
gtc Val	gag Glu 970	gcg Ala	gag Glu	gta Val	caa Gln	att Ile 975	gac Asp	agg Arg	tta Leu	att Ile	aca Thr 980	ggc Gly	aga Arg	ctt Leu	caa Gln
agc Ser 985	ctt Leu	caa Gln	acc Thr	tat Tyr	gta Val 990	aca Thr	caa Gln	caa Gln	cta Leu	atc Ile 995	agg Arg	gct Ala	gct Ala	gaa Glu	atc Ile 1000
agg Arg	gct Ala	tct Ser	gct Ala	aat Asn 1005	ctt Leu	gct Ala	gct Ala	act Thr	aaa Lys 1010	atg Met	tct Ser	gag Glu	tgt Cys	gtt Val 1015	
ctt Leu	gga Gly	caa Gln	tca Ser	aaa Lys 1020	aga Arg	gtt Val	gac Asp	ttt Phe	tgt Cys 1025	gga Gly	aag Lys	ggc Gly	tac Tyr	cac His 1030	
ctt atg	tcc tcc	ttc cca	caa gca	gcc ccg	cat ggt	ggt ggt	gtc ttc	cta ggt							

S226CAS108.ST25														
Leu	Met	Ser	Phe	Pro 1035	Gln	Ala	Ala	Pro	His 1040	Gly	Val	Val	Phe	Leu 1045
cat	gtc	acg	tat	gtg Val 1050	cca	tcc	cag	gag	agg Arg 1055	aac	ttc	acc	aca	gcg Ala 1060
His	Val	Thr	Tyr		Pro	Ser	Gln	Glu		Asn	Phe	Thr	Thr	
3268														
cca	gca	att	tgt	cat His 1065	gaa	ggc	aaa	gca	tac Tyr 1070	ttc	cct	cgt	gaa	ggt Gly 1075
Pro	Ala	Ile	Cys		Glu	Gly	Lys	Ala		Phe	Pro	Arg	Glu	
3313														
gtt	ttt	gtg	ttt	aat Asn 1080	ggc	act	tct	tgg	ttt Phe 1085	att	aca	cag	agg	aac Asn 1090
Val	Phe	Val	Phe		Gly	Thr	Ser	Trp		Ile	Thr	Gln	Arg	
3358														
ttc	ttt	tct	cca	caa Gln 1095	ata	att	act	aca	gac Asp 1100	aat	aca	ttt	gtc	tca Ser 1105
Phe	Phe	Ser	Pro		Ile	Ile	Thr	Thr		Asn	Thr	Phe	Val	
3403														
gga	aat	tgt	gat	gtc Val 1110	gtt	att	ggc	atc	att Ile 1115	aac	aac	aca	gtt	tat Tyr 1120
Gly	Asn	Cys	Asp		Val	Ile	Gly	Ile		Asn	Asn	Thr	Val	
3448														
gat	cct	ctg	caa	cct Pro 1125	gag	ctt	gac	tca	ttc Phe 1130	aaa	gaa	gag	ctg	gac Asp 1135
Asp	Pro	Leu	Gln		Glu	Leu	Asp	Ser		Lys	Glu	Glu	Leu	
3493														
aag	tac	ttc	aaa	aat Asn 1140	cat	aca	tca	cca	gat Asp 1145	gtt	gat	ctt	ggc	gac Asp 1150
Lys	Tyr	Phe	Lys		His	Thr	Ser	Pro		Val	Asp	Leu	Gly	
3538														
att	tca	ggc	att	aac Asn 1155	gct	tct	gtc	gtc	aac Asn 1160	att	caa	aaa	gaa	att Ile 1165
Ile	Ser	Gly	Ile		Ala	Ser	Val	Val		Ile	Gln	Lys	Glu	
3583														
gac	cgc	ctc	aat	gag Glu 1170	gtc	gct	aaa	aat	tta Leu 1175	aat	gaa	tca	ctc	att Ile 1180
Asp	Arg	Leu	Asn		Val	Ala	Lys	Asn		Asn	Glu	Ser	Leu	
3628														
gac	ctt	caa	gaa	ttg Leu 1185	gga	aaa	tat	gag	caa Gln 1190	tat	att	aaa	tgg	cct Pro 1195
Asp	Leu	Gln	Glu		Gly	Lys	Tyr	Glu		Tyr	Ile	Lys	Trp	
3673														
tgg	tat	gtt	tgg	ctc Leu 1200	ggc	ttc	att	gct	gga Gly 1205	cta	att	gcc	atc	gtc Val 1210
Trp	Tyr	Val	Trp		Gly	Phe	Ile	Ala		Leu	Ile	Ala	Ile	
3718														
atg	gtt	aca	atc	ttg Leu 1215	ctt	tgt	tgc	atg	act Thr 1220	agt	tgt	tgc	agt	tgc Cys 1225
Met	Val	Thr	Ile		Leu	Cys	Cys	Met		Ser	Cys	Cys	Ser	
3763														
ctc	aag	ggt	gca	tgc Cys 1230	tct	tgt	ggt	tct	tgc Cys 1235	tgc	aag	ttt	gat	gag Glu 1240
Leu	Lys	Gly	Ala		Ser	Cys	Gly	Ser		Cys	Lys	Phe	Asp	
3808														
gat	gac	tct	gag	cca Pro 1245	gtt	ctc	aag	ggt	gtc Val 1250	aaa	tta	cat	tac	aca Thr 1255
Asp	Asp	Ser	Glu		Val	Leu	Lys	Gly		Lys	Leu	His	Tyr	
3853														
taaacgaact tatggatttg tttatgagat tttttactct tggatcaatt actgcacagc														
3913														
cagtaaaaaat tgacaatgct tctcctgcaa gt														
3945														

<210> 3

<211> 1255

S226CAS108.ST25

<212> PRT

<213> CORONAVIRUS

<400> 3

Met Phe Ile Phe Leu Leu Phe Leu Thr Leu Thr Ser Gly Ser Asp Leu
 1 5 10 15

Asp Arg Cys Thr Thr Phe Asp Asp Val Gln Ala Pro Asn Tyr Thr Gln
 20 25 30

His Thr Ser Ser Met Arg Gly Val Tyr Tyr Pro Asp Glu Ile Phe Arg
 35 40 45

Ser Asp Thr Leu Tyr Leu Thr Gln Asp Leu Phe Leu Pro Phe Tyr Ser
 50 55 60

Asn Val Thr Gly Phe His Thr Ile Asn His Thr Phe Gly Asn Pro Val
 65 70 75 80

Ile Pro Phe Lys Asp Gly Ile Tyr Phe Ala Ala Thr Glu Lys Ser Asn
 85 90 95

Val Val Arg Gly Trp Val Phe Gly Ser Thr Met Asn Asn Lys Ser Gln
 100 105 110

Ser Val Ile Ile Ile Asn Asn Ser Thr Asn Val Val Ile Arg Ala Cys
 115 120 125

Asn Phe Glu Leu Cys Asp Asn Pro Phe Phe Ala Val Ser Lys Pro Met
 130 135 140

Gly Thr Gln Thr His Thr Met Ile Phe Asp Asn Ala Phe Asn Cys Thr
 145 150 155 160

Phe Glu Tyr Ile Ser Asp Ala Phe Ser Leu Asp Val Ser Glu Lys Ser
 165 170 175

Gly Asn Phe Lys His Leu Arg Glu Phe Val Phe Lys Asn Lys Asp Gly
 180 185 190

Phe Leu Tyr Val Tyr Lys Gly Tyr Gln Pro Ile Asp Val Val Arg Asp
 195 200 205

Leu Pro Ser Gly Phe Asn Thr Leu Lys Pro Ile Phe Lys Leu Pro Leu
 210 215 220

Gly Ile Asn Ile Thr Asn Phe Arg Ala Ile Leu Thr Ala Phe Ser Pro
 225 230 235 240

Ala	Gln	Asp	Ile	Trp 245	Gly	Thr	Ser	Ala	Ala 250	Ala	Tyr	Phe	Val	Gly 255	Tyr
Leu	Lys	Pro	Thr 260	Thr	Phe	Met	Leu	Lys 265	Tyr	Asp	Glu	Asn	Gly 270	Thr	Ile
Thr	Asp	Ala 275	Val	Asp	Cys	Ser	Gln 280	Asn	Pro	Leu	Ala	Glu 285	Leu	Lys	Cys
Ser	Val 290	Lys	Ser	Phe	Glu	Ile 295	Asp	Lys	Gly	Ile	Tyr 300	Gln	Thr	Ser	Asn
Phe 305	Arg	Val	Val	Pro	Ser 310	Gly	Asp	Val	Val	Arg 315	Phe	Pro	Asn	Ile	Thr 320
Asn	Leu	Cys	Pro	Phe 325	Gly	Glu	Val	Phe	Asn 330	Ala	Thr	Lys	Phe	Pro 335	Ser
Val	Tyr	Ala	Trp 340	Glu	Arg	Lys	Lys	Ile 345	Ser	Asn	Cys	Val	Ala 350	Asp	Tyr
Ser	Val	Leu 355	Tyr	Asn	Ser	Thr	Phe 360	Phe	Ser	Thr	Phe	Lys 365	Cys	Tyr	Gly
Val	Ser 370	Ala	Thr	Lys	Leu	Asn 375	Asp	Leu	Cys	Phe	Ser 380	Asn	Val	Tyr	Ala
Asp 385	Ser	Phe	Val	Val	Lys 390	Gly	Asp	Asp	Val	Arg 395	Gln	Ile	Ala	Pro	Gly 400
Gln	Thr	Gly	Val	Ile 405	Ala	Asp	Tyr	Asn	Tyr 410	Lys	Leu	Pro	Asp	Asp 415	Phe
Met	Gly	Cys	Val 420	Leu	Ala	Trp	Asn	Thr 425	Arg	Asn	Ile	Asp	Ala 430	Thr	Ser
Thr	Gly	Asn 435	Tyr	Asn	Tyr	Lys	Tyr 440	Arg	Tyr	Leu	Arg	His 445	Gly	Lys	Leu
Arg	Pro 450	Phe	Glu	Arg	Asp	Ile 455	Ser	Asn	Val	Pro	Phe 460	Ser	Pro	Asp	Gly
Lys 465	Pro	Cys	Thr	Pro	Pro 470	Ala	Leu	Asn	Cys	Tyr 475	Trp	Pro	Leu	Asn	Asp 480
Tyr	Gly	Phe	Tyr	Thr 485	Thr	Thr	Gly	Ile	Gly 490	Tyr	Gln	Pro	Tyr	Arg 495	Val
Val	Val	Leu	Ser 500	Phe	Glu	Leu	Leu	Asn 505	Ala	Pro	Ala	Thr	Val 510	Cys	Gly

S226CAS108.ST25

Pro Lys Leu Ser Thr Asp Leu Ile Lys Asn Gln Cys Val Asn Phe Asn
515 520 525

Phe Asn Gly Leu Thr Gly Thr Gly Val Leu Thr Pro Ser Ser Lys Arg
530 535 540

Phe Gln Pro Phe Gln Gln Phe Gly Arg Asp Val Ser Asp Phe Thr Asp
545 550 555 560

Ser Val Arg Asp Pro Lys Thr Ser Glu Ile Leu Asp Ile Ser Pro Cys
565 570 575

Ser Phe Gly Gly Val Ser Val Ile Thr Pro Gly Thr Asn Ala Ser Ser
580 585 590

Glu Val Ala Val Leu Tyr Gln Asp Val Asn Cys Thr Asp Val Ser Thr
595 600 605

Ala Ile His Ala Asp Gln Leu Thr Pro Ala Trp Arg Ile Tyr Ser Thr
610 615 620

Gly Asn Asn Val Phe Gln Thr Gln Ala Gly Cys Leu Ile Gly Ala Glu
625 630 635 640

His Val Asp Thr Ser Tyr Glu Cys Asp Ile Pro Ile Gly Ala Gly Ile
645 650 655

Cys Ala Ser Tyr His Thr Val Ser Leu Leu Arg Ser Thr Ser Gln Lys
660 665 670

Ser Ile Val Ala Tyr Thr Met Ser Leu Gly Ala Asp Ser Ser Ile Ala
675 680 685

Tyr Ser Asn Asn Thr Ile Ala Ile Pro Thr Asn Phe Ser Ile Ser Ile
690 695 700

Thr Thr Glu Val Met Pro Val Ser Met Ala Lys Thr Ser Val Asp Cys
705 710 715 720

Asn Met Tyr Ile Cys Gly Asp Ser Thr Glu Cys Ala Asn Leu Leu Leu
725 730 735

Gln Tyr Gly Ser Phe Cys Thr Gln Leu Asn Arg Ala Leu Ser Gly Ile
740 745 750

Ala Ala Glu Gln Asp Arg Asn Thr Arg Glu Val Phe Ala Gln Val Lys
755 760 765

Gln Met Tyr Lys Thr Pro Thr Leu Lys Tyr Phe Gly Gly Phe Asn Phe
770 775 780

S226CAS108.ST25

Ser Gln Ile Leu Pro Asp Pro Leu Lys Pro Thr Lys Arg Ser Phe Ile
785 790 795 800

Glu Asp Leu Leu Phe Asn Lys Val Thr Leu Ala Asp Ala Gly Phe Met
805 810 815

Lys Gln Tyr Gly Glu Cys Leu Gly Asp Ile Asn Ala Arg Asp Leu Ile
820 825 830

Cys Ala Gln Lys Phe Asn Gly Leu Thr Val Leu Pro Pro Leu Leu Thr
835 840 845

Asp Asp Met Ile Ala Ala Tyr Thr Ala Ala Leu Val Ser Gly Thr Ala
850 855 860

Thr Ala Gly Trp Thr Phe Gly Ala Gly Ala Ala Leu Gln Ile Pro Phe
865 870 875 880

Ala Met Gln Met Ala Tyr Arg Phe Asn Gly Ile Gly Val Thr Gln Asn
885 890 895

Val Leu Tyr Glu Asn Gln Lys Gln Ile Ala Asn Gln Phe Asn Lys Ala
900 905 910

Ile Ser Gln Ile Gln Glu Ser Leu Thr Thr Thr Ser Thr Ala Leu Gly
915 920 925

Lys Leu Gln Asp Val Val Asn Gln Asn Ala Gln Ala Leu Asn Thr Leu
930 935 940

Val Lys Gln Leu Ser Ser Asn Phe Gly Ala Ile Ser Ser Val Leu Asn
945 950 955 960

Asp Ile Leu Ser Arg Leu Asp Lys Val Glu Ala Glu Val Gln Ile Asp
965 970 975

Arg Leu Ile Thr Gly Arg Leu Gln Ser Leu Gln Thr Tyr Val Thr Gln
980 985 990

Gln Leu Ile Arg Ala Ala Glu Ile Arg Ala Ser Ala Asn Leu Ala Ala
995 1000 1005

Thr Lys Met Ser Glu Cys Val Leu Gly Gln Ser Lys Arg Val Asp
1010 1015 1020

Phe Cys Gly Lys Gly Tyr His Leu Met Ser Phe Pro Gln Ala Ala
1025 1030 1035

Pro His Gly Val Val Phe Leu His Val Thr Tyr Val Pro Ser Gln
1040 1045 1050

S226CAS108.ST25

Glu Arg Asn Phe Thr Thr Ala Pro Ala Ile Cys His Glu Gly Lys
 1055 1060 1065

Ala Tyr Phe Pro Arg Glu Gly Val Phe Val Phe Asn Gly Thr Ser
 1070 1075 1080

Trp Phe Ile Thr Gln Arg Asn Phe Phe Ser Pro Gln Ile Ile Thr
 1085 1090 1095

Thr Asp Asn Thr Phe Val Ser Gly Asn Cys Asp Val Val Ile Gly
 1100 1105 1110

Ile Ile Asn Asn Thr Val Tyr Asp Pro Leu Gln Pro Glu Leu Asp
 1115 1120 1125

Ser Phe Lys Glu Glu Leu Asp Lys Tyr Phe Lys Asn His Thr Ser
 1130 1135 1140

Pro Asp Val Asp Leu Gly Asp Ile Ser Gly Ile Asn Ala Ser Val
 1145 1150 1155

Val Asn Ile Gln Lys Glu Ile Asp Arg Leu Asn Glu Val Ala Lys
 1160 1165 1170

Asn Leu Asn Glu Ser Leu Ile Asp Leu Gln Glu Leu Gly Lys Tyr
 1175 1180 1185

Glu Gln Tyr Ile Lys Trp Pro Trp Tyr Val Trp Leu Gly Phe Ile
 1190 1195 1200

Ala Gly Leu Ile Ala Ile Val Met Val Thr Ile Leu Leu Cys Cys
 1205 1210 1215

Met Thr Ser Cys Cys Ser Cys Leu Lys Gly Ala Cys Ser Cys Gly
 1220 1225 1230

Ser Cys Cys Lys Phe Asp Glu Asp Asp Ser Glu Pro Val Leu Lys
 1235 1240 1245

Gly Val Lys Leu His Tyr Thr
 1250 1255

<210> 4

<211> 3943

<212> DNA

<213> CORONAVIRUS

<400> 4

ctcttctgga aaaaggtagg cttatcatta gagaaaacaa cagagttgtg gtttcaagtg

60

S226CAS108.ST25

atattcttgt	taacaactaa	acgaacatgt	ttattttctt	attattttctt	actctcacta	120
gtggtagtga	ccttgaccgg	tgcaccactt	ttgatgatgt	tcaagctcct	aattacactc	180
aacatacttc	atctatgagg	ggggtttact	atcctgatga	aattttttaga	tcagacactc	240
tttattttaac	tcaggattta	tttcttccat	tttattctaa	tgttacaggg	tttcatacta	300
ttaatcatac	gtttggcaac	cctgtcatac	cttttaagga	tggtatttat	tttgctgcca	360
cagagaaatc	aaatgttgtc	cgtggttggg	tttttggttc	taccatgaac	aacaagtcac	420
agtcggtgat	tattattaac	aattctacta	atgttggtat	acgagcatgt	aactttgaat	480
tgtgtgacaa	ccctttcttt	gctgtttcta	aacccatggg	tacacagaca	catactatga	540
tattcgataa	tgcatttaac	tgcactttcg	agtacatatc	tgatgccttt	tcgcttgatg	600
tttcagaaaa	gtcaggtaac	tttaaacact	tacgagagtt	tgtgtttaaa	aataaagatg	660
ggtttctcta	tgtttataag	ggctatcaac	ctatagatgt	agttcgtgat	ctaccttctg	720
gttttaacac	tttgaaacct	atttttaagt	tgcttcttgg	tattaacatt	acaaatttta	780
gagccattct	tacagccttt	tcacctgctc	aagacatttg	gggcacgtca	gctgcagcct	840
attttgttgg	ctattttaag	ccaactacat	ttatgctcaa	gtatgatgaa	aatggtacaa	900
tcacagatgc	tgttgattgt	tctcaaaatc	cacttgctga	actcaaatgc	tctgttaaga	960
gctttgagat	tgacaaagga	atttaccaga	cctctaattt	cagggttggt	ccctcaggag	1020
atgttgtagag	attccctaac	attacaaact	tgtgtccttt	tgagagaggt	tttaatgcta	1080
ctaaattccc	ttctgtctat	gcatgggaga	gaaaaaaaaa	ttctaattgt	gttgctgatt	1140
actctgtgct	ctacaactca	acattttttt	caacctttta	gtgctatggc	gtttctgcca	1200
ctaagttgaa	tgatctttgc	ttctccaatg	tctatgcaga	ttcttttgta	gtcaagggag	1260
atgatgtaag	acaaatagcg	ccaggacaaa	ctggtgttat	tgctgattat	aattataaat	1320
tgccagatga	tttcatgggt	tgtgtccttg	cttgggaatac	taggaacatt	gatgctactt	1380
caactggtaa	ttataattat	aaatataggt	atcttagaca	tggcaagctt	aggccctttg	1440
agagagacat	atctaattgt	cctttctccc	ctgatggcaa	accttgcacc	ccacctgctc	1500
ttaattgtta	ttggccatta	aatgattatg	gtttttacac	cactactggc	attggctacc	1560
aaccttacag	agttgtagta	ctttcttttg	aactttttaa	tgcaccggcc	acggtttggtg	1620
gaccaaaatt	atccactgac	cttattaaga	accagtgtgt	caattttaat	tttaatggac	1680
tcactggtac	tggtgtgtta	actccttctt	caaagagatt	tcaaccattt	caacaatttg	1740
gccgtgatgt	ctctgatttc	actgattccg	ttcgagatcc	taaaacatct	gaaatattag	1800
acatttcacc	ttgctctttt	gggggtgtaa	gtgtaattac	acctggaaca	aatgcttcat	1860
ctgaagtgc	tgttctatat	caagatgtta	actgcactga	tgtttctaca	gcaatccatg	1920
cagatcaact	cacaccagct	tggcgcatat	attctactgg	aaacaatgta	ttccagactc	1980
aagcaggctg	tcttatagga	gctgagcatg	tcgacacttc	ttatgagtgc	gacattccta	2040
ttggagctgg	catttgctgt	agttaceata	cagtttcttt	attacgtagt	actagccaaa	2100

S226CAS108.ST25

```

aatctattgt ggcttatact atgtcttttag gtgctgatag ttcaattgct tactctaata 2160
acaccattgc tatacctact aacttttcaa ttagcattac tacagaagta atgcctgttt 2220
ctatggctaa aacctccgta gattgtaata tgtacatctg cggagattct actgaatgtg 2280
ctaatttgct tctccaatat ggtagctttt gcacacaact aaatcgtgca ctctcaggta 2340
ttgctgctga acaggatcgc aacacacgtg aagtgttcgc tcaagtcaaa caaatgtaca 2400
aaacccaac tttgaaatat tttggtggtt ttaatttttc acaaataatta cctgaccctc 2460
taaagccaac taagaggtct tttattgagg acttgctctt taataagggtg acactcgtctg 2520
atgctggctt catgaagcaa tatggcgaat gcctagggtga tattaatgct agagatctca 2580
tttgtgcgca gaagttcaat gggcttacag tgttgccacc tctgctcact gatgatatga 2640
ttgctgccta cactgctgct ctagtttagtg gtactgccac tgctggatgg acatttggtg 2700
ctggcgctgc tcttcaaata ctttttgcta tgcaaatggc atatagggttc aatggcattg 2760
gagttacca aaatgttctc tatgagaacc aaaaacaaat cgccaaccaa tttacaagg 2820
cgattagtca aattcaagaa tcacttaca caacatcaac tgcattgggc aagctgcaag 2880
acgttgtaa ccagaatgct caagcattaa acacacttgt taaacaactt agctctaatt 2940
ttggtgcaat ttcaagtgtg ctaaatagata tcctttcgcg acttgataaa gtcgaggcgg 3000
aggtacaaat tgacaggcta attacaggca gacttcaaag cttcaaacc tatgtaacac 3060
aacaactaat cagggtgct gaaatcaggg cttctgctaa tcttgctgct actaaaatgt 3120
ctgagtgtgt tcttgacaa tcaaaaagag ttgacttttg tggaaagggc taccacctta 3180
tgtccttccc acaagcagcc ccgcatggtg ttgtcttcct acatgtcacg tatgtgccat 3240
cccaggagag gaacttcacc acagcgccag caatttgtca tgaaggcaaa gcatacttcc 3300
ctcgtgaagg tgttttttgtg tttaatggca cttcttggtt tattacacag aggaacttct 3360
tttctccaca aataattact acagacaata catttgtctc aggaaattgt gatgtcggtta 3420
ttggcatcat taacaacaca gtttatgatc ctctgcaacc tgagcttgac tcattcaaag 3480
aagagctgga caagtacttc aaaaatcata catcaccaga tgttgatctt ggcgacattt 3540
caggcattaa cgcttctgtc gtcaacattc aaaaagaaat tgaccgcctc aatgaggctg 3600
ctaaaaattt aaatgaatca ctcatcgacc ttcaagaatt gggaaaatat gagcaatata 3660
ttaaatggcc ttggtatgtt tggctcggct tcattgctgg actaattgcc atcgatcatg 3720
ttacaatctt gctttgttgc atgactagtt gttgcagttg cctcaagggt gcatgctctt 3780
gtggttcttg ctgcaagttt gatgaggatg actctgagcc agttctcaag ggtgtcaaat 3840
tacattacac ataaacgaac ttatggattt gtttatgaga ttttttactc ttggatcaat 3900
tactgcacag ccagtaaaaa ttgacaatgc ttctcctgca agt 3943

```

<210> 5

<211> 2049

S226CAS108.ST25

<212> DNA

<213> CORONAVIRUS

<400> 5

ctcttctgga	aaaaggtagg	cttatcatta	gagaaaacaa	cagagttgtg	gtttcaagtg	60
atattcttgt	taacaactaa	acgaacatgt	ttattttctt	attatttctt	actctcacta	120
gtggtagtga	ccttgaccgg	tgcaccactt	ttgatgatgt	tcaagctcct	aattacactc	180
aacatacttc	atctatgagg	ggggtttact	atcctgatga	aattttttaga	tcagacactc	240
tttatttaac	tcaggattta	tttcttccat	tttattctaa	tgttacaggg	tttcatacta	300
ttaatcatac	gtttggcaac	cctgtcatac	cttttaagga	tggtatttat	tttgctgcca	360
cagagaaatc	aaatgttgtc	cggtgttggg	tttttggttc	taccatgaac	aacaagtcac	420
agtcggtgat	tattattaac	aattctacta	atgttggtat	acgagcatgt	aactttgaat	480
tgtgtgacaa	ccctttcttt	gctgtttcta	aacccatggg	tacacagaca	catactatga	540
tattcgataa	tgcatttaac	tgcactttcg	agtacatata	tgatgccttt	tcgcttgatg	600
tttcagaaaa	gtcaggtaac	tttaaacact	tacgagagtt	tgtgtttaaa	aataaagatg	660
ggtttctcta	tgtttataag	ggctatcaac	ctatagatgt	agttcgtgat	ctaccttctg	720
gttttaaacac	tttgaaacct	atttttaagt	tgcctcttgg	tattaacatt	acaaatttta	780
gagccattct	tacagccttt	tcacctgctc	aagacatttg	gggcacgtca	gctgcagcct	840
attttggttg	ctattttaag	ccaactacat	ttatgctcaa	gtatgatgaa	aatggtacaa	900
tcacagatgc	tggtgattgt	tctcaaaatc	cacttgctga	actcaaatgc	tctgttaaga	960
gctttgagat	tgacaaagga	atttaccaga	cctctaattt	cagggttggt	ccctcaggag	1020
atgttggtgag	attccctaata	attacaaact	tgtgtccttt	tggagagggt	tttaatgcta	1080
ctaaattccc	ttctgtctat	gcatgggaga	gaaaaaaaaa	ttctaattgt	gttgctgatt	1140
actctgtgct	ctacaactca	acattttttt	caacctttta	gtgctatggc	gtttctgcca	1200
ctaagttgaa	tgatctttgc	ttctccaatg	tctatgcaga	ttcttttgta	gtcaagggag	1260
atgatgtaag	acaaatagcg	ccaggacaaa	ctggtgttat	tgctgattat	aattataaat	1320
tgccagatga	tttcatgggt	tgtgtccttg	cttggaaata	taggaacatt	gatgctactt	1380
caactggtaa	ttataattat	aaatataggt	atcttagaca	tggcaagctt	aggccctttg	1440
agagagacat	atctaattgt	cctttctccc	ctgatggcaa	accttgacc	ccacctgctc	1500
ttaattgtta	ttggccatta	aatgattatg	gtttttacac	cactactggc	attggctacc	1560
aaccttacag	agttgtagta	ctttcttttg	aactttttaa	tgcaccggcc	acggtttgtg	1620
gacaaaatt	atccactgac	cttattaaga	accagtgtgt	caattttaat	tttaatggac	1680
tcactggtag	tggtgtgtta	actccttctt	caaagagatt	tcaaccattt	caacaatttg	1740
gccgtgatgt	ctctgatttc	actgattccg	ttcgagatcc	taaaacatct	gaaatattag	1800
acatttcacc	ttgctctttt	gggggtgtaa	gtgtaattac	acctggaaca	aatgcttcat	1860

S226CAS108.ST25

ctgaagttgc tgttctatat caagatgtta actgcactga tgtttctaca gcaatccatg	1920
cagatcaact cacaccagct tggcgcatat attctactgg aaacaatgta ttccagactc	1980
aagcaggctg tcttatagga gctgagcatg tcgacacttc ttatgagtgc gacattccta	2040
ttggagctg	2049

<210> 6

<211> 2027

<212> DNA

<213> CORONAVIRUS

<400> 6

catgcagatc aactcacacc agcttggcgc atatattcta ctggaaacaa tgtattccag	60
actcaagcag gctgtcttat aggagctgag catgtcgaca cttcttatga gtgcgacatt	120
cctattggag ctggcatttg tgctagttag catacagttt ctttattacg tagtactagc	180
caaaaatcta ttgtggctta tactatgtct ttaggtgctg atagttcaat tgcttactct	240
aataacacca ttgctatacc tactaacttt tcaattagca ttactacaga agtaatgcct	300
gtttctatgg ctaaaacctc cgtagattgt aatatgtaca tctgcggaga ttctactgaa	360
tgtgctaatt tgcttctcca atatggtagc ttttgcacac aactaaatcg tgcactctca	420
ggatttgctg ctgaacagga tcgcaacaca cgtgaagtgt tcgctcaagt caaacaatg	480
tacaaaaccc caactttgaa atattttggg ggttttaatt tttcacaat attacctgac	540
cctctaaagc caactaagag gtcttttatt gaggacttgc tctttaataa ggtgacactc	600
gctgatgctg gcttcatgaa gcaatatggc gaatgcctag gtgatattaa tgctagagat	660
ctcatttgctg cgcagaagtt caatgggctt acagtgttgc cacctctgct cactgatgat	720
atgattgctg cctacactgc tgctctagtt agtggtagct ccactgctgg atggacattt	780
ggtgctggcg ctgctcttca aatacctttt gctatgcaaa tggcatatag gttcaatggc	840
attggaggtta cccaaaatgt tctctatgag aacccaaaac aaatcgccaa ccaatttaac	900
aaggcgatta gtcaaattca agaatcactt acaacaacat caactgcatt gggcaagctg	960
caagacgttg ttaaccagaa tgctcaagca ttaaacacac ttgttaaaca acttagctct	1020
aattttggtg caatttcaag tgtgctaaat gatatcctt cgcgacttga taaagtcgag	1080
gcgagggtac aaattgacag gttaattaca ggcagacttc aaagccttca aacctatgta	1140
acacaacaac taatcagggc tgctgaaatc agggcttctg ctaatcttgc tgctactaaa	1200
atgtctgagt gtgttcttgg acaatcaaaa agagttgact tttgtggaaa gggctaccac	1260
cttatgtcct tcccacaagc agccccgcat ggtgttgtct tcctacatgt cacgtatgtg	1320
ccatcccagg agaggaactt caccacagcg ccagcaattt gtcatgaagg caaagcatac	1380
ttccctcgtg aagggtgtttt tgtgtttaat ggcacttctt gggtttattac acagaggaa	1440

S226CAS108.ST25

```

ttcttttctc cacaaataat tactacagac aatacatttg tctcaggaaa ttgtgatgtc 1500
gttattggcg tcattaacaa cacagtattat gatcctctgc aacctgagct tgactcattc 1560
aaagaagagc tggacaagta cttcaaaaat catacatcac cagatgttga tcttggcgac 1620
atttcaggca ttaacgcttc tgtcgtcaac attcaaaaag aaattgaccg cctcaatgag 1680
gtcgttaaaa atttaaata gaatcactcatt gaccttcaag aattgggaaa atatgagcaa 1740
tatattaaat ggccttggtg tggttggttc ggcttcattg ctggactaat tgccatcgtc 1800
atggttacaa tcttgctttg ttgcatgact agttgttgca gttgcctcaa ggggtgcatgc 1860
tcttggtggt cttgctgcaa gtttgatgag gatgactctg agccagttct caagggtgtc 1920
aaattacatt acacataaac gaacttatgg atttgtttat gagatTTTTT actcttggat 1980
caattactgc acagccagta aaaattgaca atgcttctcc tgcaagt 2027

```

<210> 7

<211> 1096

<212> DNA

<213> CORONAVIRUS

<400> 7

```

tcttgctttg ttgcatgact agttgttgca gttgcctcaa ggggtgcatgc tcttggtggt 60
cttgctgcaa gtttgatgag gatgactctg agccagttct caagggtgtc aaattacatt 120
acacataaac gaacttatgg atttgtttat gagatTTTTT actcttggat caattactgc 180
acagccagta aaaattgaca atgcttctcc tgcaagtact gttcatgcta cagcaacgat 240
accgctacaa gcctcactcc ctttcggatg gcttggttatt ggcgttgcat ttcttgctgt 300
ttttcagagc gctacaaaaa taattgctgt caataaaaga tggcagctag ccctttataa 360
gggcttccag ttcatttgca atttactgct gctatttggt accatctatt cacatctttt 420
gcttgctcgt gcagggtatgg aggcgcaatt tttgtacctc tatgccttga tatattttct 480
acaatgcata aacgcatgta gaattattat gagatgttgg ctttggttga agtgcaaatc 540
caagaacca ttactttatg atgccaacta ctttggttgc tggcacacac ataactatga 600
ctactgtata ccatataaca gtgtcacaga tacaattgtc gttactgaag gtgacggcat 660
ttcaacacca aaactcaaag aagactacca aattggtggt tattctgagg ataggcactc 720
aggtgttaaa gactatgtcg ttgtacatgg ctatttcacc gaagtttact accagcttga 780
gtctacacaa attactacag aactgggtat tgaaaatgct acattcttca tctttaacaa 840
gcttggttaa gaccaccga atgtgcaaat acacacaatc gacggctctt caggagtgtc 900
taatccagca atggatccaa tttatgatga gccgacgacg actactagcg tgcctttgta 960
agcacaagaa agtgagtacg aacttatgta ctcatctgtt tcggaagaaa cagggtacgtt 1020
aatagttaat agcgtacttc tttttcttgc tttcgtggta ttcttgctag tcacactagc 1080

```

S226CAS108.ST25

catccttact gcgctt

1096

<210> 8

<211> 1135

<212> DNA

<213> CORONAVIRUS

<400> 8

attgccatcg tcatggttac aatcttgctt tgttgcacga ctagttgttg cagttgcctc	60
aagggtgcat gctcttgctg ttcttgctgc aagtttgatg aggatgactc tgagccagtt	120
ctcaagggtg tcaaattaca ttacacataa acgaacttat ggatttggtt atgagatttt	180
ttactcttggt atcaattact gcacagccag taaaaattga caatgcttct cctgcaagta	240
ctgttcacgc tacagcaacg ataccgctac aagcctcact ccctttcgga tggcttggtta	300
ttggcgcttg atttcttgct gtttttcaga gcgctaccaa aataattgcg ctcaataaaa	360
gatggcagct agccctttat aagggtctcc agttcatttg caatttactg ctgctatttg	420
ttaccatcta ttcacatctt ttgcttgctg ctgcaggtat ggaggcgcaa tttttgtacc	480
tctatgcctt gatataattt ctacaatgca tcaacgcacg tagaattatt atgagatgtt	540
ggctttgttg gaagtgcaaa tccaagaacc cattacttta tgatgccaac tactttgttt	600
gctggcacac acataactat gactactgta taccatataa cagtgtcaca gatacaattg	660
tcgttactga aggtgacggc atttcaacac caaaactcaa agaagactac caaattgggtg	720
gttattctga ggataggcac tcagggtgta aagactatgt cgttggtacat ggctatttca	780
ccgaagttta ctaccagctt gagtctacac aaattactac agacactggt attgaaaatg	840
ctacattctt catctttaac aagcttggtta aagacccacc gaatgtgcaa atacacacaa	900
tcgacggctc ttcaggagtt gctaattccag caatggatcc aatttatgat gagccgacga	960
cgactactag cgtgcctttg taagcacaag aaagtgcagta cgaacttatg tactcattcg	1020
tttcggaaga aacaggtacg ttaatagtta atagcgtact tctttttctt gctttcgtgg	1080
tattcttgct agtcacacta gccatcctta ctgcgcttcg attgtgtgcg tactg	1135

<210> 9

<211> 1096

<212> DNA

<213> CORONAVIRUS

<220>

<221> CDS

S226CAS108.ST25

<222> (137)..(958)

<223>

<400> 9

tcttgctttg	ttgcatgact	agttgttgca	gttgccctcaa	gggtgcatgc	tcttggtggtt	60
cttgctgcaa	gtttgatgag	gatgactctg	agccagttct	caagggtgtc	aaattacatt	120
acacataaac	gaactt	atg	gat	ttg	ttt	172
	Met	Asp	Leu	Phe	Met	
	1			5		
					Phe	
					Thr	
					Leu	
					Gly	
					Ser	
					10	
att	act	gca	cag	cca	gta	
Ile	Thr	Ala	Gln	Pro	Val	
		15				
					Lys	
					Ile	
					20	
					Asp	
					Asn	
					Ala	
					Ser	
					25	
					Pro	
					Ala	
					Ser	
					Thr	
					220	
gtt	cat	gct	aca	gca	acg	
Val	His	Ala	Thr	Ala	Thr	
		30				
					Ile	
					35	
					Pro	
					Leu	
					Gln	
					Ala	
					40	
					Ser	
					Leu	
					Pro	
					Phe	
					Gly	
					268	
tgg	ctt	gtt	att	ggc	gtt	
Trp	Leu	Val	Ile	Gly	Val	
		45			50	
					Ala	
					Phe	
					Leu	
					Ala	
					55	
					Val	
					Phe	
					Gln	
					Ser	
					Ala	
					Thr	
					60	
					acc	
					316	
aaa	ata	att	gcg	ctc	aat	
Lys	Ile	Ile	Ala	Leu	Asn	
				65		
					Lys	
					Arg	
					Trp	
					70	
					Gln	
					Leu	
					Ala	
					Leu	
					75	
					Tyr	
					Lys	
					Gly	
					364	
ttc	cag	ttc	att	tgc	aat	
Phe	Gln	Phe	Ile	Cys	Asn	
			80			
					Leu	
					Leu	
					85	
					Leu	
					Leu	
					Phe	
					Val	
					90	
					Thr	
					Ile	
					Tyr	
					Ser	
					412	
cat	ctt	ttg	ctt	gtc	gct	
His	Leu	Leu	Leu	Val	Ala	
		95			Ala	
					Gly	
					100	
					Met	
					Glu	
					Ala	
					Gln	
					105	
					Phe	
					Leu	
					Tyr	
					Leu	
					460	
tat	gcc	ttg	ata	tat	ttt	
Tyr	Ala	Leu	Ile	Tyr	Phe	
		110			Leu	
					Gln	
					115	
					Cys	
					Ile	
					Asn	
					Ala	
					120	
					Cys	
					Arg	
					Ile	
					Ile	
					508	
atg	aga	tgt	tgg	ctt	tgt	
Met	Arg	Cys	Trp	Leu	Cys	
				130		
					Trp	
					Lys	
					Cys	
					Lys	
					135	
					Ser	
					Lys	
					Asn	
					Pro	
					Leu	
					140	
					Leu	
					556	
tat	gat	gcc	aac	tac	ttt	
Tyr	Asp	Ala	Asn	Tyr	Phe	
				145		
					Val	
					Cys	
					Trp	
					150	
					His	
					Thr	
					His	
					Asn	
					Tyr	
					155	
					Asp	
					Tyr	
					604	
tgt	ata	cca	tat	aac	agt	
Cys	Ile	Pro	Tyr	Asn	Ser	
			160			
					Val	
					Thr	
					165	
					Asp	
					Thr	
					Ile	
					Val	
					170	
					Thr	
					Glu	
					Gly	
					652	
gac	ggc	att	tca	aca	cca	
Asp	Gly	Ile	Ser	Thr	Pro	
		175			Lys	
					Leu	
					180	
					Lys	
					Glu	
					Asp	
					Tyr	
					185	
					Gln	
					Ile	
					Gly	
					Gly	
					700	
tat	tct	gag	gat	agg	cac	
Tyr	Ser	Glu	Asp	Arg	His	
				195		
					Ser	
					Gly	
					Val	
					Lys	
					Asp	
					200	
					Tyr	
					Val	
					Val	
					Val	
					His	
					748	
ggc	tat	ttc	acc	gaa	gtt	
Gly	Tyr	Phe	Thr	Glu	Val	
				210		
					Tyr	
					Tyr	
					Gln	
					Leu	
					215	
					Glu	
					Ser	
					Thr	
					Gln	
					Ile	
					Thr	
					220	
					act	
					796	
aca	gac	act	ggc	att	gaa	
					aat	
					gct	
					aca	
					ttc	
					ttc	
					atc	
					ttt	
					aac	
					aag	
					ctt	
					844	

Leu Cys Trp Lys Cys Lys Ser Lys Asn Pro Leu Leu Tyr Asp Ala Asn
130 135 140

S226CAS108.ST25

Tyr Phe Val Cys Trp His Thr His Asn Tyr Asp Tyr Cys Ile Pro Tyr
 145 150 155 160

Asn Ser Val Thr Asp Thr Ile Val Val Thr Glu Gly Asp Gly Ile Ser
 165 170 175

Thr Pro Lys Leu Lys Glu Asp Tyr Gln Ile Gly Gly Tyr Ser Glu Asp
 180 185 190

Arg His Ser Gly Val Lys Asp Tyr Val Val Val His Gly Tyr Phe Thr
 195 200 205

Glu Val Tyr Tyr Gln Leu Glu Ser Thr Gln Ile Thr Thr Asp Thr Gly
 210 215 220

Ile Glu Asn Ala Thr Phe Phe Ile Phe Asn Lys Leu Val Lys Asp Pro
 225 230 235 240

Pro Asn Val Gln Ile His Thr Ile Asp Gly Ser Ser Gly Val Ala Asn
 245 250 255

Pro Ala Met Asp Pro Ile Tyr Asp Glu Pro Thr Thr Thr Thr Ser Val
 260 265 270

Pro Leu

<210> 11

<211> 1096

<212> DNA

<213> CORONAVIRUS

<220>

<221> CDS

<222> (558)..(1019)

<223>

<400> 11

tcttgctttg ttgcatgact agttgttgca gttgcctcaa ggggtgcatgc tcttgtggtt	60
cttgctgcaa gtttgatgag gatgactctg agccagttct caaggggtgtc aaattacatt	120
acacataaac gaacttatgg atttgtttat gagatttttt actcttggat caattactgc	180
acagccagta aaaattgaca atgcttctcc tgcaagtact gttcatgcta cagcaacgat	240
accgctacaa gcctcactcc ctttcggatg gcttggttatt ggcgttgcat ttcttgctgt	300
ttttcagagc gctaccaaaa taattgcgct caataaaaga tggcagctag ccctttataa	360

gggcttccag	ttcattttgca	atttactgct	gctatttgtt	accatctatt	cacatctttt	420										
gcttgtcgct	gcaggatatgg	aggcgcaatt	tttgtacctc	tatgccttga	tatatatttct	480										
acaatgcatc	aacgcatgta	gaattattat	gagatgttgg	ctttgttgga	agtgcaaadc	540										
caagaaccca	ttactttt	atg Met 1	atg Met 1	cca Pro	act Thr 5	act Thr 5	ttg Leu	ttt Phe	gct Ala	ggc Gly	aca Thr 10	cac His	590			
ata Ile	act Thr 15	atg Met	act Thr 15	act Thr	gta Val	tac Tyr	cat His	ata Ile 20	aca Thr	gtg Val	tca Ser	cag Gln 25	ata Ile	caa Gln	ttg Leu	638
tcg Ser	tta Leu 30	ctg Leu 30	aag Lys	gtg Val	acg Thr	gca Ala	ttt Phe 35	caa Gln	cac His	caa Gln	aac Asn 40	tca Ser	aag Lys	aag Lys	act Thr	686
acc Thr	aaa Lys 45	ttg Leu	gtg Val	gtt Val	att Ile	ctg Leu 50	agg Arg	ata Ile	ggc Gly	act Thr	cag Gln 55	gtg Val	tta Leu	aag Lys	act Thr	734
atg Met 60	tcg Ser	ttg Leu	tac Tyr	atg Met	gct Ala 65	att Ile	tca Ser	ccg Pro	aag Lys	ttt Phe 70	act Thr	acc Thr	agc Ser	ttg Leu	agt Ser 75	782
cta Leu	cac His	aaa Lys	tta Leu	cta Leu 80	cag Gln	aca Thr	ctg Leu	gta Val	ttg Leu 85	aaa Lys	atg Met	cta Leu	cat His	tct Ser 90	tca Ser	830
tct Ser	tta Leu	aca Thr	agc Ser 95	ttg Leu	tta Leu	aag Lys	acc Thr	cac His 100	cga Arg	atg Met	tgc Cys	aaa Lys	tac Tyr 105	aca Thr	caa Gln	878
tcg Ser	acg Thr	gct Ala 110	ctt Leu	cag Gln	gag Glu	ttg Leu	cta Leu 115	atc Ile	cag Gln	caa Gln	tgg Trp 120	atc Ile	caa Gln	ttt Phe	atg Met	926
atg Met	agc Ser 125	cga Arg	cga Arg	cga Arg	cta Leu	cta Leu 130	gcg Ala	tgc Cys	ctt Leu	tgt Cys	aag Lys 135	cac His	aag Lys	aaa Lys	gtg Val	974
agt Ser 140	acg Thr	aac Asn	tta Leu	tgt Cys	act Thr 145	cat His	tcg Ser	ttt Phe	cgg Arg	aag Lys 150	aaa Lys	cag Gln	gta Val	cgt Arg		1019
taatagttaa	tagcgtactt	ctttttcttg	ctttcgtggg	attcttgcta	gtcacactag	1079										
ccatccttac	tgcgctt					1096										

<213> CORONAVIRUS

Met Met Pro Thr Thr Leu Phe Ala Gly Thr His Ile Thr Met Thr Thr
1 5 10 15

S226CAS108.ST25

Val Tyr His Ile Thr Val Ser Gln Ile Gln Leu Ser Leu Leu Lys Val
 20 25 30

Thr Ala Phe Gln His Gln Asn Ser Lys Lys Thr Thr Lys Leu Val Val
 35 40 45

Ile Leu Arg Ile Gly Thr Gln Val Leu Lys Thr Met Ser Leu Tyr Met
 50 55 60

Ala Ile Ser Pro Lys Phe Thr Thr Ser Leu Ser Leu His Lys Leu Leu
 65 70 75 80

Gln Thr Leu Val Leu Lys Met Leu His Ser Ser Ser Leu Thr Ser Leu
 85 90 95

Leu Lys Thr His Arg Met Cys Lys Tyr Thr Gln Ser Thr Ala Leu Gln
 100 105 110

Glu Leu Leu Ile Gln Gln Trp Ile Gln Phe Met Met Ser Arg Arg Arg
 115 120 125

Leu Leu Ala Cys Leu Cys Lys His Lys Lys Val Ser Thr Asn Leu Cys
 130 135 140

Thr His Ser Phe Arg Lys Lys Gln Val Arg
 145 150

<210> 13

<211> 332

<212> DNA

<213> CORONAVIRUS

<220>

<221> CDS

<222> (36)..(263)

<223>

<400> 13

tgcc	tttgta	agcaca	aagaa	agtgagt	acg	aactt	atg	tac	tca	ttc	ggt	tcg		53		
							Met	Tyr	Ser	Phe	Val	Ser				
							1				5					
gaa	gaa	aca	ggt	acg	tta	ata	ggt	aat	agc	gta	ctt	ctt	ttt	ctt	gct	101
Glu	Glu	Thr	Gly	Thr	Leu	Ile	Val	Asn	Ser	Val	Leu	Leu	Phe	Leu	Ala	
			10					15					20			
ttc	gtg	gta	ttc	ttg	cta	gtc	aca	cta	gcc	atc	ctt	act	gcg	ctt	cga	149
Phe	Val	Val	Phe	Leu	Leu	Val	Thr	Leu	Ala	Ile	Leu	Thr	Ala	Leu	Arg	
			25				30					35				

S226CAS108.ST25

ttg tgt gcg tac tgc tgc aat att gtt aac gtg agt tta gta aaa cca 197
 Leu Cys Ala Tyr Cys Cys Asn Ile Val Asn Val Ser Leu Val Lys Pro
 40 45 50
 acg gtt tac gtc tac tcg cgt gtt aaa aat ctg aac tct tct gaa gga 245
 Thr Val Tyr Val Tyr Ser Arg Val Lys Asn Leu Asn Ser Ser Glu Gly
 55 60 65 70
 gtt cct gat ctt ctg gtc taaacgaact aactattatt attattctgt 293
 Val Pro Asp Leu Leu Val
 75
 ttggaacttt aacattgctt atcatggcag acaacggta 332

<210> 14

<211> 76

<212> PRT

<213> CORONAVIRUS

<400> 14

Met Tyr Ser Phe Val Ser Glu Glu Thr Gly Thr Leu Ile Val Asn Ser
 1 5 10 15
 Val Leu Leu Phe Leu Ala Phe Val Val Phe Leu Leu Val Thr Leu Ala
 20 25 30
 Ile Leu Thr Ala Leu Arg Leu Cys Ala Tyr Cys Cys Asn Ile Val Asn
 35 40 45
 Val Ser Leu Val Lys Pro Thr Val Tyr Val Tyr Ser Arg Val Lys Asn
 50 55 60
 Leu Asn Ser Ser Glu Gly Val Pro Asp Leu Leu Val
 65 70 75

<210> 15

<211> 332

<212> DNA

<213> CORONAVIRUS

<400> 15

tgcctttgta agcacaagaa agtgagtacg aacttatgta ctcatcgtt tcggaagaaa 60
 caggtacgtt aatagttaat agcgtacttc tttttcttgc tttcgtggta ttcttgctag 120
 tcacactagc catccttact gcgcttcgat tgtgtgcgta ctgctgcaat attgttaacg 180
 tgagtttagt aaaaccaacg gtttacgtct actcgcgtgt taaaaatctg aactcttctg 240
 aaggagttcc tgatcttctg gtctaaacga actaactatt attattattc tgtttggaa 300

S226CAS108.ST25

tttaacattg cttatcatgg cagacaacgg ta

332

<210> 16

<211> 708

<212> DNA

<213> CORONAVIRUS

<220>

<221> CDS

<222> (41)..(703)

<223>

<400> 16

tattattatt	attctgtttg	gaactttaac	attgcttatc	atg	gca	gac	aac	ggt	55
				Met	Ala	Asp	Asn	Gly	
				1				5	

act	att	acc	gtt	gag	gag	ctt	aaa	caa	ctc	ctg	gaa	caa	tggt	aac	cta	103
Thr	Ile	Thr	Val	Glu	Glu	Leu	Lys	Gln	Leu	Leu	Glu	Gln	Trp	Asn	Leu	
				10					15					20		

gta	ata	ggt	ttc	cta	ttc	cta	gcc	tgg	att	atg	tta	cta	caa	ttt	gcc	151
Val	Ile	Gly	Phe	Leu	Phe	Leu	Ala	Trp	Ile	Met	Leu	Leu	Gln	Phe	Ala	
			25					30					35			

tat	tct	aat	cgg	aac	agg	ttt	ttg	tac	ata	ata	aag	ctt	gtt	ttc	ctc	199
Tyr	Ser	Asn	Arg	Asn	Arg	Phe	Leu	Tyr	Ile	Ile	Lys	Leu	Val	Phe	Leu	
		40					45					50				

tgg	ctc	ttg	tgg	cca	gta	aca	ctt	gct	tgt	ttt	gtg	ctt	gct	gct	gtc	247
Trp	Leu	Leu	Trp	Pro	Val	Thr	Leu	Ala	Cys	Phe	Val	Leu	Ala	Ala	Val	
	55					60					65					

tac	aga	att	aat	tgg	gtg	act	ggc	ggg	att	gcg	att	gca	atg	gct	tgt	295
Tyr	Arg	Ile	Asn	Trp	Val	Thr	Gly	Gly	Ile	Ala	Ile	Ala	Met	Ala	Cys	
70					75					80					85	

att	gta	ggc	ttg	atg	tgg	ctt	agc	tac	ttc	gtt	gct	tcc	ttc	agg	ctg	343
Ile	Val	Gly	Leu	Met	Trp	Leu	Ser	Tyr	Phe	Val	Ala	Ser	Phe	Arg	Leu	
				90					95					100		

ttt	gct	cgt	acc	cgc	tca	atg	tgg	tca	ttc	aac	cca	gaa	aca	aac	att	391
Phe	Ala	Arg	Thr	Arg	Ser	Met	Trp	Ser	Phe	Asn	Pro	Glu	Thr	Asn	Ile	
			105					110					115			

ctt	ctc	aat	gtg	cct	ctc	cgg	ggg	aca	att	gtg	acc	aga	ccg	ctc	atg	439
Leu	Leu	Asn	Val	Pro	Leu	Arg	Gly	Thr	Ile	Val	Thr	Arg	Pro	Leu	Met	
		120					125					130				

gaa	agt	gaa	ctt	gtc	att	ggg	gct	gtg	atc	att	cgt	ggt	cac	ttg	cga	487
Glu	Ser	Glu	Leu	Val	Ile	Gly	Ala	Val	Ile	Ile	Arg	Gly	His	Leu	Arg	
	135					140					145					

atg	gcc	gga	cac	tcc	cta	ggg	cgc	tgt	gac	att	aag	gac	ctg	cca	aaa	535
Met	Ala	Gly	His	Ser	Leu	Gly	Arg	Cys	Asp	Ile	Lys	Asp	Leu	Pro	Lys	
150					155					160					165	

S226CAS108.ST25

gag atc act gtg gct aca tca cga acg ctt tct tat tac aaa tta gga 583
 Glu Ile Thr Val Ala Thr Ser Arg Thr Leu Ser Tyr Tyr Lys Leu Gly
 170 175 180

gcg tcg cag cgt gta ggc act gat tca ggt ttt gct gca tac aac cgc 631
 Ala Ser Gln Arg Val Gly Thr Asp Ser Gly Phe Ala Ala Tyr Asn Arg
 185 190 195

tac cgt att gga aac tat aaa tta aat aca gac cac gcc ggt agc aac 679
 Tyr Arg Ile Gly Asn Tyr Lys Leu Asn Thr Asp His Ala Gly Ser Asn
 200 205 210

gac aat att gct ttg cta gta cag taagt 708
 Asp Asn Ile Ala Leu Leu Val Gln
 215 220

<210> 17

<211> 221

<212> PRT

<213> CORONAVIRUS

<400> 17

Met Ala Asp Asn Gly Thr Ile Thr Val Glu Glu Leu Lys Gln Leu Leu
 1 5 10 15

Glu Gln Trp Asn Leu Val Ile Gly Phe Leu Phe Leu Ala Trp Ile Met
 20 25 30

Leu Leu Gln Phe Ala Tyr Ser Asn Arg Asn Arg Phe Leu Tyr Ile Ile
 35 40 45

Lys Leu Val Phe Leu Trp Leu Leu Trp Pro Val Thr Leu Ala Cys Phe
 50 55 60

Val Leu Ala Ala Val Tyr Arg Ile Asn Trp Val Thr Gly Gly Ile Ala
 65 70 75 80

Ile Ala Met Ala Cys Ile Val Gly Leu Met Trp Leu Ser Tyr Phe Val
 85 90 95

Ala Ser Phe Arg Leu Phe Ala Arg Thr Arg Ser Met Trp Ser Phe Asn
 100 105 110

Pro Glu Thr Asn Ile Leu Leu Asn Val Pro Leu Arg Gly Thr Ile Val
 115 120 125

Thr Arg Pro Leu Met Glu Ser Glu Leu Val Ile Gly Ala Val Ile Ile
 130 135 140

Arg Gly His Leu Arg Met Ala Gly His Ser Leu Gly Arg Cys Asp Ile
 145 150 155 160

S226CAS108.ST25

Lys Asp Leu Pro Lys Glu Ile Thr Val Ala Thr Ser Arg Thr Leu Ser
 165 170 175

Tyr Tyr Lys Leu Gly Ala Ser Gln Arg Val Gly Thr Asp Ser Gly Phe
 180 185 190

Ala Ala Tyr Asn Arg Tyr Arg Ile Gly Asn Tyr Lys Leu Asn Thr Asp
 195 200 205

His Ala Gly Ser Asn Asp Asn Ile Ala Leu Leu Val Gln
 210 215 220

<210> 18

<211> 769

<212> DNA

<213> CORONAVIRUS

<400> 18

cctgatcttc tgggtctaaac gaactaacta ttattattat tctgtttgga actttaacat	60
tgcttatcat ggcagacaac ggtactatta ccgttgagga gcttaaaca ctcctggaac	120
aatggaacct agtaataggt ttcctattcc tagcctggat tatgttacta caatttgcct	180
attctaatacg gaacagggttt ttgtacataa taaagcttgt tttcctctgg ctcttggtggc	240
cagtaacact tgcttggttt gtgcttgctg ctgtctacag aattaattgg gtgactggcg	300
ggattgcatg tgcaatggct tgtattgtag gcttgatgtg gcttagctac ttcggtgctt	360
ccttcaggct gtttgctcgt acccgctcaa tgtgggtcatt caaccagaa acaaacattc	420
ttctcaatgt gcctctccgg gggacaattg tgaccagacc gctcatggaa agtgaacttg	480
tcattgggtgc tgtgatcatt cgtgggtcact tgcgaatggc cggacactcc ctagggcgct	540
gtgacattaa ggacctgcca aaagagatca ctgtggctac atcacgaacg ctttcttatt	600
acaaattagg agcgtcgcag cgtgtaggca ctgattcagg ttttgctgca tacaaccgct	660
accgtattgg aaactataaa ttaaatacag accacgccgg tagcaacgac aatattgctt	720
tgctagtaca gtaagtgaca acagatgttt catcttggtg acttccagg	769

<210> 19

<211> 1231

<212> DNA

<213> CORONAVIRUS

<400> 19

taccgtattg gaaactataa attaaatata gaccacgccg gtagcaacga caatattgct	60
---	----

S226CAS108.ST25

```

:gctagtagtac agtaagtgtac aacagatgtt tcatcttggt gacttccagg ttacaatagc 120
agagatatattg attatcatta tgaggacttt caggattgct atttggaatc ttgacgttat 180
aataagttca atagtgtgac aattatttaa gcctctaact aagaagaatt attcggagtt 240
agatgatgaa gaacctatgg agtttagatta tccataaaac gaacatgaaa attattctct 300
tcctgacatt gattgtatatt acatcttgcg agctatatca ctatcaggag tgtgttagag 360
gtacgactgt actactaaaa gaaccttgcc catcaggaac atacgagggc aattcaccat 420
ttcacctctt tgctgacaat aaatttgac taacttgac tagcacacac tttgcttttg 480
cttggtgctga cggtagctga catacctatc agctgcgtgc aagatcagtt tcaccaaacc 540
ttttcatcag acaagaggag gttcaacaag agctctactc gccacttttt ctcatgttg 600
ctgctctagt atttttaata ctttgcttca ccattaagag aaagacagaa tgaatgagct 660
cactttaatt gacttctatt tgtgcttttt agcctttctg ctattccttg ttttaataat 720
gcttattata ttttggtttt cactcgaaat ccaggatcta gaagaacctt gtaccaaagt 780
ctaaacgaac atgaaacttc tcattgtttt gacttgtatt tctctatgca gttgcatatg 840
cactgtagta cagcgtgtg catctaataa acctcatgtg cttgaagatc cttgtaaggt 900
acaacactag gggtaatact tatagcactg cttggctttg tgctctagga aagggttttac 960
cttttcatag atggcacact atggttcaaa catgcacacc taatgttact atcaactgtc 1020
aagatccagc tgggtggtgcg cttatagcta ggtgttggtta cttcatgaa ggtcaccaaa 1080
ctgctgcatt tagagacgta cttgttggtt taaataaacg aacaaattaa aatgtctgat 1140
aatggacccc aatcaaacca acgtagtgcc cccgcatta catttggtgg acccacagat 1200
tcaactgaca ataaccagaa tggaggacgc a 1231

```

<210> 20

<211> 1242

<212> DNA

<213> CORONAVIRUS

<400> 20

```

gcatacaacc gctaccgtat tggaaactat aaattaaata cagaccacgc cggtagcaac 60
gacaatatattg ctttgctagt acagtaagtg acaacagatg tttcatcttg ttgacttcca 120
ggttacaata gcagagatat tgattatcat tatgaggact ttcaggattg ctatttgga 180
tcttgacggt ataataagtt caatagtgtg acagttatct aagcctctaa ctaagaagaa 240
ttattcggag ttagatgatg aagaacctat ggagtttagat tatccataaa acgaacatga 300
aaattattct cttcctgaca ttgattgtat ttacatcttg cgagctatat cactatcagg 360
agtgtgttag aggtacgact gtactactaa aagaaccttg cccatcagga acatacagg 420
gcaattcacc atttcaccct cttgctgaca ataaatttgc actaacttgc actagcacac 480

```

S226CAS108.ST25

```

actttgcttt tgcttgtgct gacgggtactc gacataccta tcagctgcgt gcaagatcag 540
tttcaccaaaa actttttcatc agacaagagg aggttcaaca agagctctac tcgccacttt 600
ttctcattgt tgctgctcta gtatTTTTaa tactttgctt caccattaag agaaagacag 660
aatgaatgag ctcactttaa ttgacttcta tttgtgcttt ttagcctttc tgctattcct 720
tgTTTTaata atgcttatta ttttttggtt ttcactcgaa atccaggatc tagaagaacc 780
ttgtaccaaaa gtctaaacga acatgaaact tctcattggt ttgacttgta tttctctatg 840
cagttgcata tgcactgtag tacagcgtg tgcattctaat aaacctcatg tgcttgaaga 900
tccttgtaag gtacaacact aggggtaata cttatagcac tgcttggtt tgtgctctag 960
gaaaggTTTT accttttcat agatggcaca ctatggttca aacatgcaca cctaattgta 1020
ctatcaactg tcaagatcca gctgggtggtg cgcttatagc taggtgttgg taccttcatg 1080
aaggtcacca aactgctgca tttagagacg tacttgttgt tttaaataaa cgaacgaatt 1140
aaaatgtctg ataatggacc ccaatcaaac caacgtagt ccccccgcac tacatttggt 1200
ggaccacag attcaactga caataaccag aatggaggac gc 1242

```

<210> 21

<211> 1231

<212> DNA

<213> CORONAVIRUS

<220>

<221> CDS

<222> (86)..(274)

<223>

<400> 21

```

taccgtattg gaaactataa attaaatata gaccacgccg gtagcaacga caatattgct 60
ttgctagtac agtaagtac aacag atg ttt cat ctt gtt gac ttc cag gtt 112
                               Met Phe His Leu Val Asp Phe Gln Val
                               1      5
aca ata gca gag ata ttg att atc att atg agg act ttc agg att gct 160
Thr Ile Ala Glu Ile Leu Ile Ile Ile Met Arg Thr Phe Arg Ile Ala
10      15      20      25
att tgg aat ctt gac gtt ata ata agt tca ata gtg aga caa tta ttt 208
Ile Trp Asn Leu Asp Val Ile Ile Ser Ser Ile Val Arg Gln Leu Phe
30      35      40
aag cct cta act aag aag aat tat tcg gag tta gat gat gaa gaa cct 256
Lys Pro Leu Thr Lys Lys Asn Tyr Ser Glu Leu Asp Asp Glu Glu Pro
45      50      55
atg gag tta gat tat cca taaaacgaac atgaaaatta ttctcttcct 304

```

S226CAS108.ST25

Met Glu Leu Asp Tyr Pro
60

gacattgatt gtatttacct cttgagagct atatcactat caggagtgtg ttagaggtac 364
 gactgtacta ctaaaagaac cttgcccac aggaacatac gagggcaatt caccatttca 424
 ccctcttgct gacaataaat ttgcactaac ttgcactagc acacactttg cttttgcttg 484
 tgctgacggt actcgacata cctatcagct gcgtgcaaga tcagtttcac caaaactttt 544
 catcagacaa gaggagggtc aacaagagct ctactcgcca ctttttctca ttgttgctgc 604
 tctagtattt ttaatacttt gcttcacat taagagaaag acagaatgaa tgagctcact 664
 ttaattgact tctatttgtg ctttttagcc tttctgctat tccttgtttt aataatgctt 724
 attatatttt ggttttctact cgaaatccag gatctagaag aaccttgtag caaagtctaa 784
 acgaacatga aacttctcat tgttttgact tgtatttctc tatgcagttg catatgcact 844
 gtagtacagc gctgtgcatc taataaacct catgtgcttg aagatccttg taaggtacaa 904
 cactaggggt aatacttata gcactgcttg gctttgtgct ctaggaaagg ttttaccttt 964
 tcatagatgg cacactatgg ttcaaactg cacacctaat gttactatca actgtcaaga 1024
 tccagctggt ggtgcgctta tagctaggtg ttggtacctt catgaagggt accaaactgc 1084
 tgcatttaga gacgtacttg ttgttttaaa taaacgaaca aattaaaatg tctgataatg 1144
 gacccaatc aaaccaacgt agtgccccc gcattacatt tgggtggacc acagattcaa 1204
 ctgacaataa ccagaatgga ggacgca 1231

<210> 22

<211> 63

<212> PRT

<213> CORONAVIRUS

<400> 22

Met Phe His Leu Val Asp Phe Gln Val Thr Ile Ala Glu Ile Leu Ile
1 5 10 15Ile Ile Met Arg Thr Phe Arg Ile Ala Ile Trp Asn Leu Asp Val Ile
20 25 30Ile Ser Ser Ile Val Arg Gln Leu Phe Lys Pro Leu Thr Lys Lys Asn
35 40 45Tyr Ser Glu Leu Asp Asp Glu Glu Pro Met Glu Leu Asp Tyr Pro
50 55 60

<210> 23

<211> 1231

S226CAS108.ST25

<212> DNA

<213> CORONAVIRUS

<220>

<221> CDS

<222> (285)..(650)

<223>

<400> 23

taccgtattg	gaaactataa	attaaatata	gaccacgccg	gtagcaacga	caatattgct	60
ttgctagtac	agtaagtac	aacagatggt	tcatcttggt	gacttccagg	ttacaatagc	120
agagatatgg	attatcatta	tgaggacttt	caggattgct	atttggaatc	ttgacgttat	180
aataagttca	atagtgagac	aattatttaa	gcctctaact	aagaagaatt	attcggagtt	240
agatgatgaa	gaacctatgg	agttagatta	tccataaaac	gaac atg aaa att att	296	
				Met Lys Ile Ile		
				1		

ctc ttc ctg aca ttg att gta ttt aca tct tgc gag cta tat cac tat	344
Leu Phe Leu Thr Leu Ile Val Phe Thr Ser Cys Glu Leu Tyr His Tyr	
5 10 15 20	

cag gag tgt gtt aga ggt acg act gta cta cta aaa gaa cct tgc cca	392
Gln Glu Cys Val Arg Gly Thr Thr Val Leu Leu Lys Glu Pro Cys Pro	
25 30 35	

tca gga aca tac gag ggc aat tca cca ttt cac cct ctt gct gac aat	440
Ser Gly Thr Tyr Glu Gly Asn Ser Pro Phe His Pro Leu Ala Asp Asn	
40 45 50	

aaa ttt gca cta act tgc act agc aca cac ttt gct ttt gct tgt gct	488
Lys Phe Ala Leu Thr Cys Thr Ser Thr His Phe Ala Phe Ala Cys Ala	
55 60 65	

gac ggt act cga cat acc tat cag ctg cgt gca aga tca gtt tca cca	536
Asp Gly Thr Arg His Thr Tyr Gln Leu Arg Ala Arg Ser Val Ser Pro	
70 75 80	

aaa ctt ttc atc aga caa gag gag gtt caa caa gag ctg tac tcg cca	584
Lys Leu Phe Ile Arg Gln Glu Glu Val Gln Gln Glu Leu Tyr Ser Pro	
85 90 95 100	

ctt ttt ctg att gtt gct gct cta gta ttt tta ata ctt tgc ttc acc	632
Leu Phe Leu Ile Val Ala Ala Leu Val Phe Leu Ile Leu Cys Phe Thr	
105 110 115	

att aag aga aag aca gaa tgaatgagct cactttaatt gacttctatt	680
Ile Lys Arg Lys Thr Glu	
120	

tgtgcttttt agcctttctg ctattccttg ttttaataat gcttattata ttttggtttt	740
---	-----

cactcgaaat ccaggatcta gaagaacctt gtaccaaagt cttaaactgaac atgaaacttc	800
---	-----

tcattgtttt gacttgatt tctctatgca gttgcatatg cactgtagta cagcgctgtg	860
--	-----

catctaataa acctcatgtg cttgaagatc cttgtaagggt acaacactag gggtataact	920
--	-----

S226CAS108.ST25

tatagcactg cttggctttg tgctctagga aaggttttac cttttcatag atggcacact 980
 atggttcaaa catgcacacc taatgttact atcaactgtc aagatccagc tgggtggtgcg 1040
 cttatagcta ggtgttggtg ctttcatgaa ggtcaccaaa ctgctgcatt tagagacgta 1100
 cttgttgttt taaataaacg aacaaattaa aatgtctgat aatggacccc aatcaaacca 1160
 acgtagtgcc ccccgatta catttggtgg acccacagat tcaactgaca ataaccagaa 1220
 tggaggacgc a 1231

<210> 24

<211> 122

<212> PRT

<213> CORONAVIRUS

<400> 24

Met Lys Ile Ile Leu Phe Leu Thr Leu Ile Val Phe Thr Ser Cys Glu
 1 5 10 15

Leu Tyr His Tyr Gln Glu Cys Val Arg Gly Thr Thr Val Leu Leu Lys
 20 25 30

Glu Pro Cys Pro Ser Gly Thr Tyr Glu Gly Asn Ser Pro Phe His Pro
 35 40 45

Leu Ala Asp Asn Lys Phe Ala Leu Thr Cys Thr Ser Thr His Phe Ala
 50 55 60

Phe Ala Cys Ala Asp Gly Thr Arg His Thr Tyr Gln Leu Arg Ala Arg
 65 70 75 80

Ser Val Ser Pro Lys Leu Phe Ile Arg Gln Glu Glu Val Gln Gln Glu
 85 90 95

Leu Tyr Ser Pro Leu Phe Leu Ile Val Ala Ala Leu Val Phe Leu Ile
 100 105 110

Leu Cys Phe Thr Ile Lys Arg Lys Thr Glu
 115 120

<210> 25

<211> 1231

<212> DNA

<213> CORONAVIRUS

S226CAS108.ST25

<220>

<221> CDS

<222> (650)..(781)

<223>

<400> 25

taccgtattg gaaactataa attaaatata gaccacgccg gtagcaacga caatattgct	60
ttgctagtac agtaagtgac aacagatggt tcatcttggt gacttccagg ttacaatagc	120
agagatattg attatcatta tgaggacttt caggattgct atttggaatc ttgacgttat	180
aataagttca atagttagac aattatttta gcctctaact aagaagaatt attcggaggt	240
agatgatgaa gaacctatgg agtttagatta tccataaaac gaacatgaaa attatttctt	300
tcctgacatt gattgtattt acatcttgcg agctatatca ctatcaggag tgtgttagag	360
gtacgactgt actactaaaa gaaccttgcc catcaggaac atacgagggc aattcaccat	420
ttcaccctct tgctgacaat aaatttgcac taacttgcac tagcacacac tttgcttttg	480
cttggtgctga cggtagtcga catacctatc agctgcgtgc aagatcagtt tcaccaaacc	540
ttttcatcag acaagaggag gttcaacaag agctctactc gccacttttt ctattgttg	600
ctgctctagt atttttaata ctttgcttca ccattaagag aaagacaga atg aat gag	658
	Met Asn Glu
	1
ctc act tta att gac ttc tat ttg tgc ttt tta gcc ttt ctg cta ttc	706
Leu Thr Leu Ile Asp Phe Tyr Leu Cys Phe Leu Ala Phe Leu Leu Phe	
5 10 15	
ctt gtt tta ata atg ctt att ata ttt tgg ttt tca ctc gaa atc cag	754
Leu Val Leu Ile Met Leu Ile Ile Phe Trp Phe Ser Leu Glu Ile Gln	
20 25 30 35	
gat cta gaa gaa cct tgt acc aaa gtc taaacgaaca tgaaacttct	801
Asp Leu Glu Glu Pro Cys Thr Lys Val	
40	
cattgttttg acttgatatt ctctatgcag ttgcatatgc actgtagtac agcgctgtgc	861
atctaataaa cctcatgtgc ttgaagatcc ttgtaaggta caacactagg ggtaataactt	921
atagcactgc ttggctttgt gctctaggaa aggttttacc ttttcataga tggcacacta	981
tggttcaaac atgcacacct aatgttacta tcaactgtca agatccagct ggtggtgcgc	1041
ttatagctag gtgttggtac cttcatgaag gtcaccaaac tgctgcattt agagacgtac	1101
ttgttgtttt aaataaacga acaaattaaa atgtctgata atggacccca atcaaaccac	1161
cgtagtgcc cccgcattac atttggtgga cccacagatt caactgacaa taaccagaat	1221
ggaggacgca	1231

<210> 26

<211> 44

S226CAS108.ST25

<212> PRT

<213> CORONAVIRUS

<400> 26

Met Asn Glu Leu Thr Leu Ile Asp Phe Tyr Leu Cys Phe Leu Ala Phe
 1 5 10 15

Leu Leu Phe Leu Val Leu Ile Met Leu Ile Ile Phe Trp Phe Ser Leu
 20 25 30

Glu Ile Gln Asp Leu Glu Glu Pro Cys Thr Lys Val
 35 40

<210> 27

<211> 1231

<212> DNA

<213> CORONAVIRUS

<220>

<221> CDS

<222> (791)..(907)

<223>

<400> 27
 taccgtattg gaaactataa attaaatata gaccacgccg gtagcaacga caatattgct 60
 ttgctagtac agtaagtgaac aacagatggt tcattctgtt gacttccagg ttacaatagc 120
 agagatattg attatcatta tgaggacttt caggattgct atttggaatc ttgacgttat 180
 aataagttca atagtgaagac aattatttta gcctctaact aagaagaatt attcggagtt 240
 agatgatgaa gaacctatgg agttagatta tccataaaac gaacatgaaa attattctct 300
 tcctgacatt gattgtattt acatcttgcg agctatatca ctatcaggag tgtgttagag 360
 gtacgactgt actactaaaa gaaccttgcc catcaggaac atacgagggc aattcaccat 420
 ttcaccctct tgctgacaat aaatttgcac taacttgcac tagcacacac tttgcttttg 480
 cttgtgctga cggtagctga catacctatc agctgcgtgc aagatcagtt tcaccaaacc 540
 ttttcatcag acaagaggag gttcaacaag agctctactc gccacttttt ctcatgtgtg 600
 ctgctctagt atttttaata ctttgcttca ccattaagag aaagacagaa tgaatgagct 660
 cactttaatt gacttctatt tgtgcttttt agcctttctg ctattccttg ttttaataat 720
 gcttattata ttttggtttt cactcgaaat ccaggatcta gaagaacctt gtaccaaagt 780
 ctaaacgaac atg aaa ctt ctc att gtt ttg act tgt att tct cta tgc 829

S226CAS108.ST25

Met Lys Leu Leu Ile Val Leu Thr Cys Ile Ser Leu Cys
 1 5 10

agt tgc ata tgc act gta gta cag cgc tgt gca tct aat aaa cct cat 877
 Ser Cys Ile Cys Thr Val Val Gln Arg Cys Ala Ser Asn Lys Pro His
 15 20 25

gtg ctt gaa gat cct tgt aag gta caa cac taggggtaat acttatagca 927
 Val Leu Glu Asp Pro Cys Lys Val Gln His
 30 35

ctgcttggct ttgtgctcta ggaaagggtt taccttttca tagatggcac actatgggttc 987
 aaacatgcac acctaattgtt actatcaact gtcaagatcc agctggtggt gcgcttatag 1047
 ctaggtgttg gtaccttcat gaaggtcacc aaactgctgc atttagagac gtacttggtg 1107
 ttttaaataa acgaacaaat taaaatgtct gataatggac cccaatcaaa ccaacgtagt 1167
 gcccccgca ttacatttgg tggaccaca gattcaactg acaataacca gaatggagga 1227
 cgca 1231

<210> 28

<211> 39

<212> PRT

<213> CORONAVIRUS

<400> 28

Met Lys Leu Leu Ile Val Leu Thr Cys Ile Ser Leu Cys Ser Cys Ile
 1 5 10 15

Cys Thr Val Val Gln Arg Cys Ala Ser Asn Lys Pro His Val Leu Glu
 20 25 30

Asp Pro Cys Lys Val Gln His
 35

<210> 29

<211> 1231

<212> DNA

<213> CORONAVIRUS

<220>

<221> CDS

<222> (876)..(1127)

<223>

S226CASI08.S123																
<400>	29	taccgtattg	gaaactataa	attaaatata	gaccacgccg	gtagcaacga	caatatttgct									60
ttgctagtag	agtaagtga	aacagatgtt	tcatcttggt	gacttccagg	ttacaatagc											120
agagatat	attatcatta	tgaggacttt	caggattgct	atttggaatc	ttgacgttat											180
aataagttca	atagtga	aattat	gcctctaact	aagaagaatt	attcggagtt											240
agatgatgaa	gaacctatgg	agttagatta	tccataaaac	gaacatgaaa	attattctct											300
tcctgacatt	gattgtat	acatcttg	agctatatca	ctatcaggag	tgtgttagag											360
gtacgactgt	actactaaaa	gaaccttgcc	catcaggaac	atacgagggc	aattcaccat											420
ttcaccctct	tgctgacaat	aaatttgcac	taacttgcac	tagcacacac	tttgcttttg											480
cttgtgctga	cggtactcga	catacctatc	agctgcgtgc	aagatcagtt	tcacccaaaac											540
ttttcatcag	acaagaggag	gttcaacaag	agctctactc	gccacttttt	ctcattgttg											600
ctgctctagt	at	ctttgcttca	ccattaagag	aaagacagaa	tgaatgagct											660
cactttaatt	gacttctatt	tgtgcttttt	agcctttctg	ctattccttg	ttttaataat											720
gcttattata	ttttggtttt	cactcgaaat	ccaggatcta	gaagaacctt	gtacccaaagt											780
ctaaacgaac	atgaaacttc	tcattgtttt	gacttgtatt	tctctatgca	gttgcatatg											840
cactgtagta	cagcgctgtg	catctaataa	acctc	atg	tgc	ttg	aag	atc	ctt							893
				Met	Cys	Leu	Lys	Ile	Leu							
				1				5								
gta	agg	tac	aac	act	agg	ggg	aat	act	tat	agc	act	gct	tgg	ctt	tgt	941
Val	Arg	Tyr	Asn	Thr	Arg	Gly	Asn	Thr	Tyr	Ser	Thr	Ala	Trp	Leu	Cys	
			10					15					20			
gct	cta	gga	aag	gtt	tta	cct	ttt	cat	aga	tgg	cac	act	atg	gtt	caa	989
Ala	Leu	Gly	Lys	Val	Leu	Pro	Phe	His	Arg	Trp	His	Thr	Met	Val	Gln	
		25					30					35				
aca	tgc	aca	cct	aat	gtt	act	atc	aac	tgt	caa	gat	cca	gct	ggt	ggt	1037
Thr	Cys	Thr	Pro	Asn	Val	Thr	Ile	Asn	Cys	Gln	Asp	Pro	Ala	Gly	Gly	
	40					45					50					
gcg	ctt	ata	gct	agg	tgt	tgg	tac	ctt	cat	gaa	ggg	cac	caa	act	gct	1085
Ala	Leu	Ile	Ala	Arg	Cys	Trp	Tyr	Leu	His	Glu	Gly	His	Gln	Thr	Ala	
55					60					65					70	
gca	ttt	aga	gac	gta	ctt	gtt	gtt	tta	aat	aaa	cga	aca	aat			1127
Ala	Phe	Arg	Asp	Val	Leu	Val	Val	Leu	Asn	Lys	Arg	Thr	Asn			
				75					80							
taaaatgtct	gataatggac	cccaatcaaa	ccaacgtagt	gccccccgca	ttacatttgg											

<213> CORONAVIRUS

S226CAS108.ST25

<400> 30

Met Cys Leu Lys Ile Leu Val Arg Tyr Asn Thr Arg Gly Asn Thr Tyr
 1 5 10 15

Ser Thr Ala Trp Leu Cys Ala Leu Gly Lys Val Leu Pro Phe His Arg
 20 25 30

Trp His Thr Met Val Gln Thr Cys Thr Pro Asn Val Thr Ile Asn Cys
 35 40 45

Gln Asp Pro Ala Gly Gly Ala Leu Ile Ala Arg Cys Trp Tyr Leu His
 50 55 60

Glu Gly His Gln Thr Ala Ala Phe Arg Asp Val Leu Val Val Leu Asn
 65 70 75 80

Lys Arg Thr Asn

<210> 31

<211> 21221

<212> DNA

<213> CORONAVIRUS

<400> 31

atggagagcc ttgttcttgg tgtcaacgag aaaacacacg tccaactcag tttgcctgtc	60
cttcagggtta gagacgtgct agtgcggtggc ttcgggggact ctgtggaaga ggccctatcg	120
gaggcacgtg aacacctcaa aaatggcact tgtgggtctag tagagctgga aaaaggcgta	180
ctgccccagc ttgaacagcc ctatgtgttc attaaacggt ctgatgcctt aagcaccaat	240
cacggccaca aggtcggtga gctggttgca gaaatggacg gcattcagta cggtcgtagc	300
ggtataacac tgggagtact cgtgccacat gtgggcgaaa cccaattgc ataccgcaat	360
gttcttcttc gtaagaacgg taataaggga gccgggtggc atagctatgg catcgatcta	420
aagtcttatg acttaggtga cgagcttggc actgatccca ttgaagatta tgaacaaaac	480
tggaaacta agcatggcag tgggtgactc cgtgaactca ctcgtgagct caatggaggt	540
gcagtcactc gctatgtcga caacaatttc tgtggcccag atgggtaccc tcttgattgc	600
atcaaagatt ttctcgacg cgcgggcaag tcaatgtgca ctctttccga acaacttgat	660
tacatcgagt cgaagagagg tgtctactgc tgccgtgacc atgagcatga aattgcctgg	720
ttcactgagc gctctgataa gagctacgag caccagacac ccttcgaaat taagagtgcc	780
aagaaatttg acactttcaa aggggaatgc ccaaagtttg tgtttctct taactcaaaa	840
gtcaaagtca ttcaaccacg tgttgaaaag aaaaagactg agggtttcat ggggcgtata	900
cgctctgtgt accctgttgc atctccacag gagtgtaaca atatgcactt gtctaccttg	960

S226CAS108.ST25

atgaaatgta	atcattgcga	tgaagtttca	tggcagacgt	gcgactttct	gaaagccact	1020
tgtgaacatt	gtggcactga	aaatttagtt	attgaaggac	ctactacatg	tgggtaccta	1080
cctactaatg	ctgtagtgaa	aatgccatgt	cctgcctgtc	aagaccaga	gattggacct	1140
gagcatagtg	ttgcagatta	tcacaaccac	tcaaacattg	aaactcgact	ccgcaagggg	1200
ggtaggacta	gatgttttgg	aggctgtgtg	tttgcctatg	ttggctgcta	taataagcgt	1260
gcctactggg	ttcctcgtgc	tagtgctgat	attggctcag	gccatactgg	cattactggt	1320
gacaatgtgg	agaccttgaa	tgaggatctc	cttgagatac	tgagtcgtga	acgtgttaac	1380
attaacattg	ttggcgattt	tcatttgaat	gaagagggtg	ccatcatttt	ggcatctttc	1440
tctgcttcta	caagtgcctt	tattgacact	ataaagagtc	ttgattacaa	gtctttcaaa	1500
accattgttg	agtcctgcgg	taactataaa	gttaccaagg	gaaagcccgt	aaaagggtgct	1560
tgggaacattg	gacaacagag	atcagtttta	acaccactgt	gtggttttcc	ctcacaggct	1620
gctggtgtta	tcagatcaat	ttttgcgcgc	acacttgatg	cagcaaacca	ctcaattcct	1680
gatttgcaaa	gagcagctgt	caccatactt	gatgggtattt	ctgaacagtc	attacgtctt	1740
gtcgacgcca	tggttttatac	ttcagacctg	ctcaccaaca	gtgtcattat	tatggcatat	1800
gtaactggtg	gtcttgtaca	acagacttct	cagtgggtgt	ctaattcttt	gggcactact	1860
gttgaaaaac	tcaggcctat	ctttgaatgg	attgaggcga	aacttagtgc	aggagttgaa	1920
tttctcaagg	atgcttgggg	gatttctcaa	tttctcatta	cagggtgttt	tgacatcgtc	1980
aagggtcaaa	tacaggttgc	ttcagataac	atcaaggatt	gtgtaaaatg	cttcattgat	2040
gttgtaaca	aggcactcga	aatgtgcatt	gatcaagtca	ctatcgctgg	cgcaaagttg	2100
cgatcactca	acttaggtga	agtcttcatc	gctcaaagca	agggacttta	ccgtcagtgt	2160
atacgtggca	aggagcagct	gcaactactc	atgcctctta	aggcaccaaa	agaagtaacc	2220
tttcttgaag	gtgattcaca	tgacacagta	cttacctctg	aggaggttgt	tctcaagaac	2280
ggtgaactcg	aagcactcga	gacgcccgtt	gatagcttca	caaattggagc	tatcgttggc	2340
acaccagtct	gtgtaaatgg	cctcatgctc	ttagagatta	aggacaaaga	acaatactgc	2400
gcattgtctc	ctggtttact	ggctacaaac	aatgtctttc	gcttaaaagg	gggtgcacca	2460
attaaagggtg	taacctttgg	agaagatact	gtttgggaag	ttcaagggtta	caagaatgtg	2520
agaatcacat	ttgagcttga	tgaacgtgtt	gacaaagtgc	ttaatgaaaa	gtgctctgtc	2580
tacactgttg	aatccggtac	cgaagttact	gagtttgc	gtgtttagc	agaggctgtt	2640
gtgaagactt	tacaaccagt	ttctgatctc	cttaccaaca	tgggtattga	tcttgatgag	2700
tggagtgtag	ctacattcta	cttatttgat	gatgctgggtg	aagaaaactt	ttcatcacgt	2760
atgtattgtt	cctttttacc	tccagatgag	gaagaagagg	acgatgcaga	gtgtgaggaa	2820
gaagaaattg	atgaaacctg	tgaacatgag	tacgggtacag	aggatgatta	tcaaggctctc	2880
cctctggaat	ttggtgcctc	agctgaaaca	gttcgagttg	aggaagaaga	agaggaagac	2940
tggctggatg	atactactga	gcaatcagag	attgagccag	aaccagaacc	tacacctgaa	3000

S226CAS108.ST25

gaaccagtta atcagttttac tggttatttta aaacttactg acaatgttgc cattaaatgt	3060
gttgacatcg ttaaggagggc acaaagtgtc aatcctatgg tgattgtaaa tgctgctaac	3120
atacacctga aacatggtgg tgggtgtagca ggtgcactca acaaggcaac caatggtgcc	3180
atgcaaaagg agagtgatga ttacattaag ctaaattggcc ctcttacagt aggaggggtct	3240
tgtttgcttt ctggacataa tcttgctaag aagtgtctgc atgttggtgg acctaaccta	3300
aatgcagggtg aggacatcca gcttcttaag gcagcatatg aaaatttcaa ttcacaggac	3360
atcttacttg caccattggt gtcagcaggc atatttggtg ctaaaccact tcagtcttta	3420
caagtgtgcg tgcagacggt tcgtacacag gtttatattg cagtcaatga caaagctctt	3480
tatgagcagg ttgtcatgga ttatcttgat aacctgaagc ctagagtggga agcacctaaa	3540
caagaggagc caccaaacac agaagattcc aaaactgagg agaaatctgt cgtacagaag	3600
cctgtcgatg tgaagccaaa aattaaggcc tgcattgatg aggttaccac aacactggaa	3660
gaaactaagt ttcttaccaa taagttactc ttgtttgctg atatcaatgg taagctttac	3720
catgattctc agaacatgct tagagggtgaa gatatgtctt tccttgagaa ggatgcacct	3780
tacatggtag gtgatgttat cactagtggg gatatcactt gtgttgtaat accctccaaa	3840
aaggctgggtg gcactactga gatgctctca agagctttga agaaagtgcc agttgatgag	3900
tatataacca cgtaccctgg acaaggatgt gctgggttata cacttgagga agctaagact	3960
gctcttaaga aatgcaaadc tgcattttat gtactacctt cagaagcacc taatgctaag	4020
gaagagattc taggaactgt atcctggaat ttgagagaaa tgcttgctca tgctgaagag	4080
acaagaaaat taatgcctat atgcatggat gttagagcca taatggcaac catccaacgt	4140
aagtataaag gaattaaaat tcaagagggc atcgttgact atgggtgtccg attcttcttt	4200
tatactagta aagagcctgt agcttctatt attacgaagc tgaactctct aaatgagccg	4260
cttgtcacia tgccaattgg ttatgtgaca catgggtttta atcttgaaga ggctgcgcgc	4320
tgtatgcgtt ctcttaaagc tcctgccgta gtgtcagtat catcaccaga tgctgttact	4380
acataaatg gatacctcac ttcgtcatca aagacatctg aggagcactt tgtagaaaca	4440
gtttcttttg ctggctctta cagagattgg tcctattcag gacagcgtac agagttagggt	4500
gttgaatttc ttaagcgtgg tgacaaaatt gtgtaccaca ctctggagag ccccgtcgag	4560
tttcatcttg acggtgaggt tctttcactt gacaaactaa agagtctctt atccctgcgg	4620
gaggttaaga ctataaaagt gttcacaact gtggacaaca ctaatctcca cacacagctt	4680
gtggatatgt ctatgacata tggacagcag tttgggtcaa catacttgga tgggtgctgat	4740
gttacaaaaa ttaaacctca tgtaaactcat gagggttaaga ctttctttgt actacctagt	4800
gatgacacac tacgtagtga agctttcgag tactaccata ctcttgatga gagttttctt	4860
ggtaggtaca tgtctgcttt aaaccacaca aagaaatgga aatttcctca agttgggtgg	4920
ttaacttcaa ttaaatgggc tgataacaat tgttatttgt ctagtgtttt attagcactt	4980
caacagcttg aagtcaaatt caatgcacca gcacttcaag aggcttatta tagagcccgt	5040

S226CAS108.ST25

gctggtgatg	ctgctaactt	ttgtgcactc	atactcgctt	acagtaataa	aactgttggc	5100
gagcttggtg	atgtcagaga	aactatgacc	catcttctac	agcatgctaa	tttggaatct	5160
gcaaagcgag	ttcttaatgt	ggtgtgtaaa	cattgtggtc	agaaaactac	taccttaacg	5220
ggtgtagaag	ctgtgatgta	tatgggtact	ctatcttatg	ataatcttaa	gacagggtgtt	5280
tccattccat	gtgtgtgtgg	tcgtgatgct	acacaatatc	tagtacaaca	agagtcttct	5340
tttgttatga	tgtctgcacc	acctgctgag	tataaattac	agcaaggtag	attcttatgt	5400
gcgaatgagt	acactggtaa	ctatcagtgt	ggtcattaca	ctcatataac	tgctaaggag	5460
accctctatc	gtattgacgg	agctcacctt	acaaagatgt	cagagtacaa	aggaccagtg	5520
actgatgttt	tctacaagga	aacatcttac	actacaacca	tcaagcctgt	gtcgtataaa	5580
ctcgatggag	ttacttacac	agagattgaa	ccaaaattgg	atgggtatta	taaaaaggat	5640
aatgcttact	atacagagca	gcctatagac	cttgtagcaa	ctcaaccatt	accaaattgag	5700
agttttgata	atttcaaact	cacatgttct	aacacaaaat	ttgctgatga	tttaaataca	5760
atgacaggct	tcacaaagcc	agcttcacga	gagctatctg	tcacattctt	cccagacttg	5820
aatggcgatg	tagtggctat	tgactataga	cactattcag	cgagtttcaa	gaaagggtgct	5880
aaattactgc	ataagccaat	tgtttggcac	attaaccagg	ctacaaccaa	gacaacgttc	5940
aaaccaaaca	cttgggtgtt	acgttgtctt	tggagtacaa	agccagtaga	tacttcaaata	6000
tcatttgaag	ttctggcagt	agaagacaca	caaggaatgg	acaatcttgc	ttgtgaaagt	6060
caacaaccca	cctctgaaga	agtagtgga	aatcctacca	tacagaagga	agtcatagag	6120
tgtgacgtga	aaactaccga	agttgtaggc	aatgtcatac	ttaaaccatc	agatgaagggt	6180
gttaaagtaa	cacaagagtt	aggtcatgag	gatcttatgg	ctgcttatgt	ggaaaacaca	6240
agcattacca	ttaagaaacc	taatgagctt	tcactagcct	taggttttaa	aacaattgcc	6300
actcatggta	ttgctgcaat	taatagtgtt	ccttggagta	aaattttggc	ttatgtcaaa	6360
ccattcttag	gacaagcagc	aattacaaca	tcaaattgag	ctaagagatt	agcacaacgt	6420
gtgtttaaca	attatatgcc	ttatgtgttt	acattattgt	tccaattgtg	tacttttact	6480
aaaagtacca	attctagaat	tagagcttca	ctacctacaa	ctattgctaa	aaatagtgtt	6540
aagagtgttg	ctaaattatg	tttggatgcc	ggcattaatt	atgtgaagtc	acccaaattt	6600
tctaaattgt	tcacaatcgc	tatgtggcta	ttgttggtta	gtatttgctt	aggttctcta	6660
atctgtgtaa	ctgctgcttt	tggtgtactc	ttatctaatt	ttggtgctcc	ttcttattgt	6720
aatggcggtta	gagaattgta	tcttaattcg	tctaacgtta	ctactatgga	tttctgtgaa	6780
ggttcttttc	cttgcagcat	ttgtttaagt	ggattagact	cccttgattc	ttatccagct	6840
cttgaaacca	ttcagggtgac	gatttcatcg	tacaagctag	acttgacaat	tttaggtctg	6900
gccgctgagt	gggttttggc	atatatgttg	ttcacaaaat	tcttttattt	attaggtctt	6960
tcagctataa	tgcagggtgtt	ctttggctat	tttgctagtc	atttcatcag	caattcttgg	7020
ctcatgtggt	ttatcattag	tattgtacaa	atggcacccg	tttctgcaat	ggttaggatg	7080

S226CAS108.ST25

tacatcttct	ttgcttcttt	ctactacata	tggaagagct	atgttcatat	catggatggg	7140
tgcacctctt	cgacttgcac	gatgtgctat	aagcgcaatc	gtgccacacg	cgttgagtgt	7200
acaactattg	ttaatggcat	gaagagatct	ttctatgtct	atgcaaattg	aggccgtggc	7260
ttctgcaaga	ctcacaattg	gaattgtctc	aattgtgaca	cattttgcac	tggtagtaca	7320
ttcattagt	atgaagtgtc	tcgtgatttg	tcactccagt	ttaaaagacc	aatcaaccct	7380
actgaccagt	catcgtatat	tgttgatagt	gttgctgtga	aaaatggcgc	gcttcacctc	7440
tactttgaca	aggctgggtc	aaagacctat	gagagacatc	cgctctccca	ttttgtcaat	7500
ttagacaatt	tgagagctaa	caacactaaa	ggttcactgc	ctattaatgt	catagttttt	7560
gatggcaagt	ccaaatgcga	cgagtctgct	tctaagtctg	cttctgtgta	ctacagtcag	7620
ctgatgtgcc	aacctattct	gttgcttgac	caagctcttg	tatcagacgt	tgagatagt	7680
actgaagttt	ccgttaagat	gtttgatgct	tatgtcgaca	ccttttcagc	aacttttagt	7740
gttcctatgg	aaaaacttaa	ggcacttggt	gctacagctc	acagcgagtt	agcaaagggg	7800
gtagcttttag	atgggtgtcct	ttctacattc	gtgtcagctg	cccgacaagg	tggtgttgat	7860
accgatgttg	acacaaagga	tgttattgaa	tgtctcaaac	tttcacatca	ctctgactta	7920
gaagtgcag	gtgacagttg	taacaatttc	atgctcacct	ataataaggt	tgaaaacatg	7980
acgcccagag	atcttggcgc	atgtattgac	tgtaatgcaa	ggcatatcaa	tgcccaagta	8040
gcaaaaagtc	acaatgtttc	actcatctgg	aatgtaaaag	actacatgtc	tttatctgaa	8100
cagctgcgta	aacaaattcg	tagtgctgcc	aagaagaaca	acataccttt	tagactaact	8160
tgtgctacaa	ctagacaggt	tgtcaatgtc	ataactacta	aaatctcact	caagggtggg	8220
aagattgtta	gtacttgttt	taaacttatg	cttaaggcca	cattattgtg	cgttcttgct	8280
gcattggttt	gttatatcgt	tatgccagta	catacattgt	caatccatga	tggttacaca	8340
aatgaaatca	ttgggttacaa	agccattcag	gatgggtgtc	ctcgtgacat	catttctact	8400
gatgattgtt	ttgcaaataa	acatgctggg	tttgacgcat	ggtttagcca	gcgtgggtgg	8460
tcatacaaaa	atgacaaaag	ctgccctgta	gtagctgcta	tcattacaag	agagattggg	8520
ttcatagtgc	ctggcttacc	gggtactgtg	ctgagagcaa	tcaatgggtga	cttcttgcat	8580
tttctacctc	gtgttttttag	tgctgttggc	aacatttgct	acacaccttc	caaactcatt	8640
gagtatagt	atcttgctac	ctctgcttgc	gttcttgctg	ctgagtgtac	aatttttaag	8700
gatgctatgg	gcaaacctgt	gccatattgt	tatgacacta	atcttgctaga	gggttctatt	8760
tcttatagt	agcttcgtcc	agacactcgt	tatgtgctta	tggtatgggtc	catcatacag	8820
tttcctaaca	cttacctgga	gggttctgtt	agagtagtaa	caacttttga	tgctgagtac	8880
tgtagacatg	gtacatgcga	aaggtcagaa	gtaggtatct	gcctatctac	cagtggtaga	8940
tgggttctta	ataatgagca	ttacagagct	ctatcaggag	ttttctgtgg	tggtgatgctg	9000
atgaatctca	tagctaaccat	ctttactcct	cttgtgcaac	ctgtgggtgc	tttagatgtg	9060
tctgcttcag	tagtggtctg	tggtattatt	gccatattgg	tgacttgtgc	tgccctactac	9120

S226CAS108.ST25

tttatgaaat tcagacgtgt ttttggtgag tacaaccatg ttgttgctgc taatgcactt	9180
ttgtttttga tgtctttcac tatactctgt ctggtaccag cttacagctt tctgccggga	9240
gtctactcag tcttttactt gtacttgaca ttctatttca ccaatgatgt ttcattcttg	9300
gctcaccttc aatgggttgc catgttttct cctattgtgc ctttttggat aacagcaatc	9360
tatgtattct gtatttctct gaagcactgc cattggttct ttaacaacta tcttaggaaa	9420
agagtcatgt ttaatggagt tacatttagt accttcgagg aggctgcttt gtgtaccttt	9480
ttgctcaaca aggaaatgta cctaaaattg cgtagcgaga cactgttgcc acttacacag	9540
tataacaggt atcttgctct atataacaag tacaagtatt tcagtggagc cttagatact	9600
accagctatc gtgaagcagc ttgctgccac ttagcaaagg ctctaaatga ctttagcaac	9660
tcagggtgctg atgttctcta ccaaccacca cagacatcaa tcacttctgc tgttctgcag	9720
agtggtttta ggaaaatggc attcccgtca ggcaaagttg aagggtgcat ggtacaagta	9780
acctgtggaa ctacaactct taatggattg tggttggatg acacagtata ctgtccaaga	9840
catgtcattt gcacagcaga agacatgctt aatcctaact atgaagatct gctcattcgc	9900
aatccaacc atagctttct tgttcaggct ggcaatgttc aacttcgtgt tattggccat	9960
tctatgcaaa attgtctgct taggcttaaa gttgatactt ctaaccctaa gacaccaag	10020
tataaatttg tccgtatcca acctggtcaa acattttcag ttctagcatg ctacaatggt	10080
tcaccatctg gtgtttatca gtgtgccatg agacctaatc ataccattaa aggttctttc	10140
cttaatggat catgtggtag tgttggtttt aacattgatt atgattgcgt gtctttctgc	10200
tatatgcatc atatggagct tccaacagga gtacacgctg gtactgactt agaaggtaaa	10260
ttctatggtc catttgttga cagacaaact gcacaggctg caggtagaga cacaaccata	10320
acattaaatg ttttggtcatg gctgtatgct gctgttatca atggtgatag gtggtttctt	10380
aatagattca ccactacttt gaatgacttt aaccttgtgg caatgaagta caactatgaa	10440
cctttgacac aagatcatgt tgacatattg ggacctctt ctgctcaaac aggaattgcc	10500
gtcttagata tgtgtgctgc tttgaaagag ctgctgcaga atggtatgaa tggctgtact	10560
atccttggtg gcactatttt agaagatgag ttacaccat ttgatgttgt tagacaatgc	10620
tctggtgtta ccttccaagg taagttcaag aaaattgtta agggcactca tcattggatg	10680
cttttaactt tcttgacatc actattgatt ctgttcaaa gtacacagtg gtcactgttt	10740
ttctttgttt acgagaatgc tttcttgcca ttactcttg gtattatggc aattgctgca	10800
tgtgctatgc tgcttggtta gcataagcac gcattcttgt gcttgtttct gttaccttct	10860
cttgcaacag ttgcttactt taatatggtc tacatgcctg ctagctgggt gatgcgtatc	10920
atgacatggc ttgaattggc tgacactagc ttgtctgggt ataggcttaa ggattgtgtt	10980
atgtatgctt cagctttagt tttgcttatt ctcatgacag ctcgcactgt ttatgatgat	11040
gctgctagac gtgtttggac actgatgaat gtcattacac ttgtttacaa agtctactat	11100
ggtaatgctt tagatcaagc tatttccatg tgggccttag ttatttctgt aaccttaac	11160

S226CAS108.ST25

tattctggtg	tcgttacgac	tatcatgttt	ttagctagag	ctatagtgtt	tgtgtgtgtt	11220
gagtattacc	cattgttatt	tattactggc	aacaccttac	agtgtatcat	gcttgtttat	11280
tgtttcttag	gctattgttg	ctgctgctac	tttggccttt	tctgtttact	caaccgttac	11340
ttcaggctta	ctcttggtgt	ttatgactac	ttggtctcta	cacaagaatt	taggtatatg	11400
aactcccagg	ggcttttgcc	tcctaagagt	agtattgatg	ctttcaagct	taacattaag	11460
ttgttgggta	ttggaggtaa	accatgtatc	aagggttgcta	ctgtacagtc	taaaatgtct	11520
gacgtaaagt	gcacatctgt	ggtactgctc	tcggttcttc	aacaacttag	agtagagtca	11580
tcttctaaat	tgtgggcaca	atgtgtacaa	ctccacaatg	atattcttct	tgcaaaagac	11640
acaactgaag	ctttcgagaa	gatggtttct	cttttgctctg	ttttgctatc	catgcagggt	11700
gctgtagaca	ttaatagggt	gtgcgaggaa	atgctcgata	accgtgctac	tcttcaggct	11760
attgcttcag	aatttagttc	tttaccatca	tatgccgctt	atgccactgc	ccaggaggcc	11820
tatgagcagg	ctgtagctaa	tggtgattct	gaagtcgttc	tcaaaaagtt	aaagaaatct	11880
ttgaatgtgg	ctaaatctga	gtttgaccgt	gatgctgcca	tgcaacgcaa	gttggaaaag	11940
atggcagatc	aggctatgac	ccaaatgtac	aaacaggcaa	gatctgagga	caagagggca	12000
aaagtaacta	gtgctatgca	aacaatgctc	ttcactatgc	ttaggaagct	tgataatgat	12060
gcacttaaca	acattatcaa	caatgcgcgt	gatggttgtg	ttccactcaa	catcatacca	12120
ttgactacag	cagccaaact	catggttggt	gtccctgatt	atggtaccta	caagaacact	12180
tgtgatggta	acacctttac	atatgcatct	gcactctggg	aaatccagca	agttgttgat	12240
gcggatagca	agattgttca	acttagtgaa	attaacatgg	acaattcacc	aaatttggt	12300
tggcctctta	ttgttacagc	tctaagagcc	aactcagctg	ttaaactaca	gaataatgaa	12360
ctgagtccag	tagcactacg	acagatgtcc	tgtgcggtg	gtaccacaca	aacagcttgt	12420
actgatgaca	atgcacttgc	ctactataac	aattcgaagg	gaggtagggt	tgtgctggca	12480
ttactatcag	accaccaaga	tctcaaattg	gctagattcc	ctaagagtga	tggtacaggt	12540
acaatttaca	cagaactgga	accaccttgt	aggtttgtta	cagacacacc	aaaagggcct	12600
aaagtgaat	acttgactt	catcaaaggc	ttaaacaacc	taaatagagg	tatggtgctg	12660
ggcagtttag	ctgctacagt	acgtcttcag	gctggaaatg	ctacagaagt	acctgccaat	12720
tcaactgtgc	tttccttctg	tgcttttgca	gtagaccctg	ctaaagcata	taaggattac	12780
ctagcaagtg	gaggacaacc	aatcaccaac	tgtgtgaaga	tgttgtgtac	acacactggt	12840
acaggacagg	caattactgt	aacaccagaa	gctaacatgg	accaagagtc	ctttgggtggt	12900
gcttcatgtt	gtctgtattg	tagatgccac	attgaccatc	caaatcctaa	aggattctgt	12960
gacttgaaag	gtaagtacgt	ccaaatacct	accacttgtg	ctaatgaccc	agtgggtttt	13020
acacttagaa	acacagtctg	taccgtctgc	ggaatgtgga	aagggttatgg	ctgtagttgt	13080
gaccaactcc	gcgaaccctt	gatgcagtct	gcggatgcat	caacgttttt	aaacgggttt	13140
gcggtgtaag	tgcagcccgt	cttacaccgt	gcggcacagg	cactagtact	gatgtcgtct	13200

S226CAS108.ST25

acagggcttt tgatatttac aacgaaaaag ttgctggttt tgcaaagttc ctaaaaacta 13260
 attgctgtcg cttccaggag aaggatgagg aaggcaattt attagactct tactttgtag 13320
 ttaagaggca tactatgtct aactaccaac atgaagagac tatttataac ttggttaaag 13380
 attgtccagc gggtgctgtc catgactttt tcaagtttag agtagatggg gacatggtag 13440
 cacatatatc acgtcagcgt ctaactaaat acacaatggc tgatttagtc tatgctctac 13500
 gtcattttga tgagggtaat tgtgatacat taaaagaaat actcgtcaca tacaattgct 13560
 gtgatgatga ttatttcaat aagaaggatt ggtatgactt cgtagagaat cctgacatct 13620
 tacgcgtata tgctaactta ggtgagcgtg tacgccaatc attattaaag actgtacaat 13680
 tctgcgatgc tatgcgtgat gcaggcattg taggcgtact gacattagat aatcaggatc 13740
 ttaatgggaa ctggtacgat ttcggtgatt tcgtacaagt agcaccaggc tgcggagttc 13800
 ctattgtgga ttcataattac tcattgctga tgcccacct cactttgact agggcattgg 13860
 ctgctgagtc ccatatggat gctgatctcg caaaaccact tattaagtgg gatttgctga 13920
 aatatgattt tacggaagag agactttgtc tcttcgaccg ttattttaaa tattgggacc 13980
 agacatacca tccaattgt attaactgtt tggatgatag gtgtatcctt cattgtgcaa 14040
 actttaatgt gttattttct actgtgtttc cacctacaag ttttggacca ctagtaagaa 14100
 aaatatattt agatgggtgt cttttgtttg tttcaactgg ataccatttt cgtgagtttag 14160
 gagtcgtaca taatcaggat gtaaacttac atagctcgcg tctcagtttc aaggaaactt 14220
 tagtgatgac tgctgatcca gctatgcatg cagcttctgg caatttattg ctagataaac 14280
 gcactacatg cttttcagta gctgcactaa caaacaatgt tgcttttcaa actgtcaaac 14340
 ccggtaatth taataaagac ttttatgact ttgctgtgtc taaaggtttc ttaagggaag 14400
 gaagttctgt tgaactaaaa cacttcttct ttgctcagga tggcaacgct gctatcagtg 14460
 attatgacta ttatcgttat aatctgcaa caatgtgtga tatcagacaa ctctatttcg 14520
 tagttgaagt tggtgataaa tactttgatt gttacgatgg tggctgtatt aatgccaacc 14580
 aagtaatcgt taacaatctg gataaatcag ctggtttccc atttaataaa tggggtaagg 14640
 ctagacttta ttatgactca atgagttatg aggatcaaga tgcacttttc gcgtatacta 14700
 agcgtaatgt catccctact ataactcaa tgaatcttaa gtatgccatt agtgcaaaga 14760
 atagagctcg caccgtagct ggtgtctcta tctgtagtac tatgacaaat agacagtttc 14820
 atcagaaatt attgaagtca atagccgcca ctagaggagc tactgtggta attggaacaa 14880
 gcaagtttta cgggtggctgg cataatatgt taaaactgt ttacagtgat gtagaaactc 14940
 cacaccttat ggggtgggat tatccaaaat gtgacagagc catgcctaac atgcttagga 15000
 taatggcctc tcttgttctt gctcgcaaac ataactttg ctgtaactta tcacaccggt 15060
 tctacagggt agctaacgag tgtgcgcaag tattaagtga gatggcatg tgtggcggct 15120
 cactatatgt taaaccagggt ggaacatcat ccggtgatgc tacaactgct tatgctaata 15180
 gtgtctttaa catttgctaa gctgttacag ccaatgtaaa tgcacttctt tcaactgatg 15240

S226CAS108.ST25

gtaataagat agctgacaag tatgtccgca atctacaaca caggctctat gagtgtctct 15300
 atagaaatag ggatgttgat catgaattcg tggatgagtt ttacgcttac ctgcgtaaac 15360
 atttctccat gatgattcct tctgatgatg ccgttggtgtg ctataacagt aactatgcgg 15420
 ctcaagggtt agtagctagc attaagaact ttaaggcagt tctttattat caaaataatg 15480
 tgttcatgtc tgaggcaaaa tgttggtactg agactgacct tactaaagga cctcacgaat 15540
 ttgtgtcaca gcatacaatg ctagttaaac aaggagatga ttacgtgtac ctgccttacc 15600
 cagatccatc aagaatatta ggcgcaggct gttttgtcga tgatattgtc aaaacagatg 15660
 gtacacttat gattgaaagg ttcgtgtcac tggctattga tgcttaccac cttacaaaac 15720
 atcctaatac ggagtatgct gatgtctttc acttgatttt acaatacatt agaaagtac 15780
 atgatgagct tactggccac atgttggaca tgtattccgt aatgctaact aatgataaca 15840
 cctcacggta ctgggaacct gagttttatg aggctatgta cacaccacat acagtcttgc 15900
 aggctgtagg tgcttgtgta ttgtgcaatt cacagacttc acttcgttgc ggtgcctgta 15960
 ttaggagacc attcctatgt tgcaagtgtc gctatgacca tgtcatttca acatcacaca 16020
 aattagtgtt gtctgttaat ccctatgttt gcaatgcccc aggttgtgat gtcactgatg 16080
 tgacacaact gtatctagga ggtatgagct attattgcaa gtcacataag cctcccatta 16140
 gttttccatt atgtgctaatt ggtcagggtt ttggtttata caaaaacaca tgtgtaggca 16200
 gtgacaatgt cactgacttc aatgcgatag caacatgtga ttggactaat gctggcgatt 16260
 acatacttgc caacacttgt actgagagac tcaagctttt cgcagcagaa acgctcaaag 16320
 ccactgagga aacatttaag ctgtcatatg gtattgccac tgtacgcgaa gtactctctg 16380
 acagagaatt gcatctttca tgggaggttg gaaaacctag accaccattg aacagaaact 16440
 atgtctttac tggttaccgt gtaactaaaa atagtaaagt acagattgga gagtacacct 16500
 ttgaaaaagg tgactatggt gatgctgttg tgtacagagg tactacgaca tacaagttga 16560
 atgttgggtga ttactttgtg ttgacatctc acactgtaat gccacttagt gcacctactc 16620
 tagtgccaca agagcactat gtgagaatta ctggcttgta cccaacactc aacatctcag 16680
 atgagttttc tagcaatgtt gcaaattatc aaaaggctcg catgcaaaag tactctacac 16740
 tccaaggacc acctggtact ggtaagagtc attttgccat cggacttgct ctctattacc 16800
 catctgctcg catagtgtat acggcatgct ctcatgcagc tgttgatgcc ctatgtgaaa 16860
 aggcattaaa atatttgccc atagataaat gtagtagaat catacctgcg cgtgcgcgcg 16920
 tagagtgttt tgataaattc aaagtgaatt caacactaga acagtatgtt ttctgcactg 16980
 taaatgcatt gccagaaaca actgctgaca ttgtagtctt tgatgaaatc tctatggcta 17040
 ctaattatga cttgagtgtt gtcaatgcta gacttcgtgc aaaacactac gtctatattg 17100
 gcgatcctgc tcaattacca gcccccgca cattgctgac taaaggcaca ctagaaccag 17160
 aatattttta ttcagtgtgc agacttatga aaacaatagg tccagacatg ttccttgga 17220
 cttgtcgccg ttgtcctgct gaaattgttg aactgtgag tgcttttagt tatgacaata 17280

S226CAS108.ST25

agctaaaagc acacaaggat aagtcagctc aatgcttcaa aatgttctac aaaggtgtta 17340
 ttacacatga tgtttcatct gcaatcaaca gacctcaaat aggcgttgta agagaatttc 17400
 ttacacgcaa tcctgcttg agaaaagctg tttttatctc accttataat tcacagaacg 17460
 ctgtagcttc aaaaatctta ggattgccta cgcagactgt tgattcatca cagggttctg 17520
 aatatgacta tgcataattc acacaaacta ctgaaacagc acactcttgt aatgtcaacc 17580
 gcttcaatgt ggctatcaca agggcaaaaa ttggcatttt gtgcataatg tctgatagag 17640
 atctttatga caaactgcaa tttaacaagtc tagaaatacc acgtcgcaat gtggctacat 17700
 tacaagcaga aaatgtaact ggacttttta aggactgtag taagatcatt actggtcttc 17760
 atcctacaca ggcacctaca cacctcagcg ttgatataaa gttcaagact gaaggattat 17820
 gtgttgacat accaggcata ccaaaggaca tgacctaccg tagactcatc tctatgatgg 17880
 gtttcaaaat gaattaccaa gtcaatgggt accctaatat gtttatcacc cgcgaagaag 17940
 ctattcgtca cgttcgtgcg tggattggct ttgatgtaga gggctgtcat gcaactagag 18000
 atgctgtggg tactaaccta cctctccagc taggattttc tacagggtgtt aacttagtag 18060
 ctgtaccgac tggttatggt gacactgaaa ataacacaga attcaccaga gttaatgcaa 18120
 aacctccacc aggtgaccag tttaaacatc ttataccact catgtataaa ggcttgccct 18180
 ggaatgtagt gcgtattaag atagtacaaa tgctcagtga tacactgaaa ggattgtcag 18240
 acagagtcgt gttcgtcctt tgggcgcatg gctttgagct tacatcaatg aagtactttg 18300
 tcaagattgg acctgaaaga acgtgttgct tgtgtgacaa acgtgcaact tgcttttcta 18360
 cttcatcaga tacttatgcc tgctggaatc attctgtggg ttttgactat gtctataacc 18420
 ctttatgat tgatgttcag cagtggggct ttacgggtaa cttcagagt aaccatgacc 18480
 aacattgcca ggtacatgga aatgcacatg tggctagtgt tgatgctatc atgactagat 18540
 gtttagcagt ccatgagtgc tttgttaagc gcgttgattg gtctgttgaa taccctatta 18600
 taggagatga actgagggtt aattctgctt gcagaaaagt acaacacatg gttgtgaagt 18660
 ctgcattgct tgctgataag tttccagttc ttcattgacat tggaaatcca aaggctatca 18720
 agtgtgtgcc tcaggctgaa gtagaatgga agttctacga tgctcagcca tgtagtgaca 18780
 aagcttaca aatagaggaa ctcttctatt cttatgctac acatcacgat aaattcactg 18840
 atgggtgtttg tttgttttgg aattgtaacg ttgatcgta cccagccaat gcaattgtgt 18900
 gtaggtttga cacaagagtc ttgtcaaact tgaacttacc aggctgtgat ggtggtagtt 18960
 tgtatgtgaa taagcatgca ttccacactc cagctttcga taaaagtgca tttactaatt 19020
 taaagcaatt gcctttcttt tactattctg atagtccttg tgagtctcat ggcaaacaag 19080
 tagtgtcgga tattgattat gttccactca aatctgctac gtgtattaca cgatgcaatt 19140
 taggtggtgc tgtttgacaga caccatgcaa atgagtaccg acagtacttg gatgcatata 19200
 atatgatgat ttctgctgga tttagcctat ggatttaca acaatttgat acttataacc 19260
 tgtggaatac atttaccagg ttacagagtt tagaaaatgt ggcttataat gttgttaata 19320

S226CAS108.ST25

```

aaggacactt tgatggacac gccggcgaag cacctgtttc catcattaat aatgctgttt 19380
acacaaaggt agatgggtatt gatgtggaga tctttgaaaa taagacaaca cttcctgtta 19440
atgttgcatt tgagcttttg gctaagcgta acattaaacc agtgccagag attaagatac 19500
tcaataattht ggggtgttgat atcgttgcta atactgtaat ctgggactac aaaagagaag 19560
ccccagcaca tgtatctaca ataggtgtct gcacaatgac tgacattgcc aagaaaccta 19620
ctgagagtgc ttgttcttca cttactgtct tgtttgatgg tagagtggaa ggacaggtag 19680
accttttttag aaacgcccgt aatgggtgttt taataacaga aggttcagtc aaagggtctaa 19740
caccttcaaa gggaccagca caagctagcg tcaatggagt cacattaatt ggagaatcag 19800
taaaaaacaca gttaactac ttaagaaag tagacggcat tattcaacag ttgcctgaaa 19860
cctactttac tcagagcaga gacttagagg attttaagcc cagatcaca atggaaactg 19920
actttctcga gctcgtatg gatgaattca tacagcgata taagctcgag ggctatgcct 19980
tcgaacacat cgtttatgga gatttcagtc atggacaact tggcggctctt catttaatga 20040
taggcttagc caagcgctca caagattcac cacttaaatt agaggatttt atccctatgg 20100
acagcacagt gaaaaattac ttcataacag atgcgcaaac aggttcatca aaatgtgtgt 20160
gttctgtgat tgatctttta cttgatgact ttgtcgagat aataaagtca caagatttgt 20220
cagtgtattc aaaagtggc aaggttacaa ttgactatgc tgaaatttca ttcattgttt 20280
ggtgtaagga tggacatgtt gaaaccttct acccaaaact acaagcaagt caagcgtggc 20340
aaccagggtg tgcgatgcct aacttgatca agatgcaaag aatgcttctt gaaaagtgtg 20400
accttcagaa ttatggtgaa aatgctgtta taccaaaagg aataatgatg aatgtcgcaa 20460
agtatactca actgtgtcaa tacttaata cacttacttt agctgtaccc tacaacatga 20520
gagttattca ctttggtgct ggctctgata aaggagttgc accagggtaca gctgtgtcga 20580
gacaatgggt gccaaactggc acactacttg tcgattcaga tcttaatgac ttcgtctccg 20640
acgcagattc tacttttaatt ggagactgtg caacagtaca tacggctaataaatgggacc 20700
ttattattag cgatatgtat gaccctagga ccaaacatgt gacaaaagag aatgactcta 20760
aagaaggggt tttcacttat ctgtgtggat ttataaagca aaaactagcc ctgggtgggt 20820
ctatagctgt aaagataaca gagcattctt ggaatgctga cttttacaag cttatgggcc 20880
atttctcatg gtggacagct tttgttaca atgtaaatgc atcatcatcg gaagcatttt 20940
taattggggc taactatctt ggcaagccga aggaacaaat tgatggctat accatgcatg 21000
ctaactacat tttctggagg aacacaaatc ctatccagtt gtcttcctat tcaactcttg 21060
acatgagcaa atttcctctt aaattaagag gaactgctgt aatgtctctt aaggagaatc 21120
aatcaatga tatgatttat tctcttctgg aaaaaggtag gcttatcatt agagaaaaca 21180
acagagttgt ggtttcaagt gatattcttg ttaacaacta a 21221

```


S226CAS108.ST25

<211> 297

<212> DNA

<213> CORONAVIRUS

<400> 32

```

atggaccca atcaaacc aa cgtagtgccc cccgcattac atttggtgga cccacagatt      60
caactgacaa taaccagaat ggaggacgca atggggcaag gccaaaacag cgccgacccc      120
aaggtttacc caataatact gcgtcttggt tcacagctct cactcagcat ggcaaggagg      180
aacttagatt ccctcgaggc cagggcgttc caatcaacac caatagtggg ccagatgacc      240
aaattggcta ctaccgaaga gctacccgac gagttcgtgg tggtgacggc aaaatga      297

```

<210> 33

<211> 98

<212> PRT

<213> CORONAVIRUS

<400> 33

```

Met Asp Pro Asn Gln Thr Asn Val Val Pro Pro Ala Leu His Leu Val
1      5      10      15
Asp Pro Gln Ile Gln Leu Thr Ile Thr Arg Met Glu Asp Ala Met Gly
20      25      30
Gln Gly Gln Asn Ser Ala Asp Pro Lys Val Tyr Pro Ile Ile Leu Arg
35      40      45
Leu Gly Ser Gln Leu Ser Leu Ser Met Ala Arg Arg Asn Leu Asp Ser
50      55      60
Leu Glu Ala Arg Ala Phe Gln Ser Thr Pro Ile Val Val Gln Met Thr
65      70      75      80
Lys Leu Ala Thr Thr Glu Glu Leu Pro Asp Glu Phe Val Val Val Thr
85      90      95
Ala Lys

```

<210> 34

<211> 213

<212> DNA

<213> CORONAVIRUS

S226CAS108.ST25

<400> 34
 atgctgccac cgtgctacaa cttcctcaag gaacaacatt gccaaaaggc ttctacgcag 60
 agggaagcag aggcggcagt caagcctctt ctcgctcctc atcacgtagt cgcggtaatt 120
 caagaaattc aactcctggc agcagtaggg gaaattctcc tgctcgaatg gctagcggag 180
 gtggtgaaac tgccctcgcg ctattgctgc tag 213

<210> 35
 <211> 70
 <212> PRT
 <213> CORONAVIRUS

<400> 35
 Met Leu Pro Pro Cys Tyr Asn Phe Leu Lys Glu Gln His Cys Gln Lys
 1 5 10 15
 Ala Ser Thr Gln Arg Glu Ala Glu Ala Val Lys Pro Leu Leu Ala
 20 25 30
 Pro His His Val Val Ala Val Ile Gln Glu Ile Gln Leu Leu Ala Ala
 35 40 45
 Val Gly Glu Ile Leu Leu Leu Glu Trp Leu Ala Glu Val Val Lys Leu
 50 55 60
 Pro Ser Arg Tyr Cys Cys
 65 70

<210> 36
 <211> 1377
 <212> DNA
 <213> CORONAVIRUS

<220>
 <221> CDS
 <222> (67)..(1335)
 <223>

<400> 36
 atgaaggcca ccaaactgct gcatttagag acgtacttgt tgttttaaataaacgaacaa 60
 attaaa atg tct gat aat gga ccc caa tca aac caa cgt agt gcc ccc 108

S226CAS108.ST25

Met Ser Asp Asn Gly Pro Gln Ser Asn Gln Arg Ser Ala Pro
 1 5 10

cgc	att	aca	ttt	ggt	gga	ccc	aca	gat	tca	act	gac	aat	aac	cag	aat	156
Arg	Ile	Thr	Phe	Gly	Gly	Pro	Thr	Asp	Ser	Thr	Asp	Asn	Asn	Gln	Asn	
15				20				25						30		
gga	gga	cgc	aat	ggg	gca	agg	cca	aaa	cag	cgc	cga	ccc	caa	ggt	tta	204
Gly	Gly	Arg	Asn	Gly	Ala	Arg	Pro	Lys	Gln	Arg	Arg	Pro	Gln	Gly	Leu	
				35				40						45		
ccc	aat	aat	act	gcg	tct	tgg	ttc	aca	gct	ctc	act	cag	cat	ggc	aag	252
Pro	Asn	Asn	Thr	Ala	Ser	Trp	Phe	Thr	Ala	Leu	Thr	Gln	His	Gly	Lys	
			50				55						60			
gag	gaa	ctt	aga	ttc	cct	cga	ggc	cag	ggc	gtt	cca	atc	aac	acc	aat	300
Glu	Glu	Leu	Arg	Phe	Pro	Arg	Gly	Gln	Gly	Val	Pro	Ile	Asn	Thr	Asn	
		65					70					75				
agt	ggt	cca	gat	gac	caa	att	ggc	tac	tac	cga	aga	gct	acc	cga	cga	348
Ser	Gly	Pro	Asp	Asp	Gln	Ile	Gly	Tyr	Tyr	Arg	Arg	Ala	Thr	Arg	Arg	
	80					85					90					
gtt	cgt	ggt	ggt	gac	ggc	aaa	atg	aaa	gag	ctc	agc	ccc	aga	tgg	tac	396
Val	Arg	Gly	Gly	Asp	Gly	Lys	Met	Lys	Glu	Leu	Ser	Pro	Arg	Trp	Tyr	
95				100					105					110		
ttc	tat	tac	cta	gga	act	ggc	cca	gaa	gct	tca	ctt	ccc	tac	ggc	gct	444
Phe	Tyr	Tyr	Leu	Gly	Thr	Gly	Pro	Glu	Ala	Ser	Leu	Pro	Tyr	Gly	Ala	
				115				120						125		
aac	aaa	gaa	ggc	atc	gta	tgg	gtt	gca	act	gag	gga	gcc	ttg	aat	aca	492
Asn	Lys	Glu	Gly	Ile	Val	Trp	Val	Ala	Thr	Glu	Gly	Ala	Leu	Asn	Thr	
			130					135					140			
ccc	aaa	gac	cac	att	ggc	acc	cgc	aat	cct	aat	aac	aat	gct	gcc	acc	540
Pro	Lys	Asp	His	Ile	Gly	Thr	Arg	Asn	Pro	Asn	Asn	Asn	Ala	Ala	Thr	
		145					150					155				
gtg	cta	caa	ctt	cct	caa	gga	aca	aca	ttg	cca	aaa	ggc	ttc	tac	gca	588
Val	Leu	Gln	Leu	Pro	Gln	Gly	Thr	Thr	Leu	Pro	Lys	Gly	Phe	Tyr	Ala	
	160					165					170					
gag	gga	agc	aga	ggc	ggc	agt	caa	gcc	tct	tct	cgc	tcc	tca	tca	cgt	636
Glu	Gly	Ser	Arg	Gly	Gly	Ser	Gln	Ala	Ser	Ser	Arg	Ser	Ser	Ser	Arg	
175				180					185						190	
agt	cgc	ggt	aat	tca	aga	aat	tca	act	cct	ggc	agc	agt	agg	gga	aat	684
Ser	Arg	Gly	Asn	Ser	Arg	Asn	Ser	Thr	Pro	Gly	Ser	Ser	Arg	Gly	Asn	
				195					200					205		
tct	cct	gct	cga	atg	gct	agc	gga	ggt	ggt	gaa	act	gcc	ctc	gcg	cta	732
Ser	Pro	Ala	Arg	Met	Ala	Ser	Gly	Gly	Gly	Glu	Thr	Ala	Leu	Ala	Leu	
			210				215						220			
ttg	ctg	cta	gac	aga	ttg	aac	cag	ctt	gag	agc	aaa	gtt	tct	ggt	aaa	780
Leu	Leu	Leu	Asp	Arg	Leu	Asn	Gln	Leu	Glu	Ser	Lys	Val	Ser	Gly	Lys	
		225					230					235				
ggc	caa	caa	caa	caa	ggc	caa	act	gtc	act	aag	aaa	tct	gct	gct	gag	828
Gly	Gln	Gln	Gln	Gln	Gly	Gln	Thr	Val	Thr	Lys	Lys	Ser	Ala	Ala	Glu	
	240					245					250					
gca	tct	aaa	aag	cct	cgc	caa	aaa	cgt	act	gcc	aca	aaa	cag	tac	aac	876
Ala	Ser	Lys	Lys	Pro	Arg	Gln	Lys	Arg	Thr	Ala	Thr	Lys	Gln	Tyr	Asn	
255				260					265						270	
gtc	act	caa	gca	ttt	ggg	aga	cgt	ggt	cca	gaa	caa	acc	caa	gga	aat	924

S226CAS108.ST25

val	Thr	Gln	Ala	Phe 275	Gly	Arg	Arg	Gly	Pro 280	Glu	Gln	Thr	Gln	Gly 285	Asn		
ttc	ggg	gac	caa	gac	cta	atc	aga	caa	gga	act	gat	tac	aaa	cat	tgg	972	
Phe	Gly	Asp	Gln 290	Asp	Leu	Ile	Arg	Gln 295	Gly	Thr	Asp	Tyr	Lys 300	His	Trp		
ccg	caa	att	gca	caa	ttt	gct	cca	agt	gcc	tct	gca	ttc	ttt	gga	atg	1020	
Pro	Gln	Ile 305	Ala	Gln	Phe	Ala	Pro 310	Ser	Ala	Ser	Ala	Phe 315	Phe	Gly	Met		
tca	cgc	att	ggc	atg	gaa	gtc	aca	cct	tcg	gga	aca	tgg	ctg	act	tat	1068	
Ser	Arg	Ile 320	Gly	Met	Glu	Val 325	Thr	Pro	Ser	Gly	Thr 330	Trp	Leu	Thr	Tyr		
cat	gga	gcc	att	aaa	ttg	gat	gac	aaa	gat	cca	caa	ttc	aaa	gac	aac	1116	
His	Gly	Ala	Ile	Lys	Leu 340	Asp	Asp	Lys	Asp	Pro 345	Gln	Phe	Lys	Asp	Asn 350		
gtc	ata	ctg	ctg	aac	aag	cac	att	gac	gca	tac	aaa	aca	ttc	cca	cca	1164	
Val	Ile	Leu	Leu	Asn 355	Lys	His	Ile	Asp	Ala 360	Tyr	Lys	Thr	Phe	Pro 365	Pro		
aca	gag	cct	aaa	aag	gac	aaa	aag	aaa	aag	act	gat	gaa	gct	cag	cct	1212	
Thr	Glu	Pro	Lys 370	Lys	Asp	Lys	Lys	Lys 375	Lys	Thr	Asp	Glu	Ala 380	Gln	Pro		
ttg	ccg	cag	aga	caa	aag	aag	cag	ccc	act	gtg	act	ctt	ctt	cct	gcg	1260	
Leu	Pro	Gln 385	Arg	Gln	Lys	Lys	Gln 390	Pro	Thr	Val	Thr 395	Leu	Leu	Pro	Ala		
gct	gac	atg	gat	gat	ttc	tcc	aga	caa	ctt	caa	aat	tcc	atg	agt	gga	1308	
Ala	Asp	Met	Asp	Asp	Phe	Ser 405	Arg	Gln	Leu	Gln	Asn 410	Ser	Met	Ser	Gly		
gct	tct	gct	gat	tca	act	cag	gca	taa	acact	catga	tgaccacaca					1355	
Ala	Ser	Ala	Asp	Ser	Thr 420	Gln	Ala										
415																	
aggcagatgg	gctatgtaaa	cg														1377	

<210> 37

<211> 422

<212> PRT

<213> CORONAVIRUS

<400> 37

Met Ser Asp Asn Gly Pro Gln Ser Asn Gln Arg Ser Ala Pro Arg Ile
1 5 10 15

Thr Phe Gly Gly Pro Thr Asp Ser Thr Asp Asn Asn Gln Asn Gly Gly
20 25 30

Arg Asn Gly Ala Arg Pro Lys Gln Arg Arg Pro Gln Gly Leu Pro Asn
35 40 45

Asn Thr Ala Ser Trp Phe Thr Ala Leu Thr Gln His Gly Lys Glu Glu
50 55 60

S226CAS108.ST25

Leu Arg Phe Pro Arg Gly Gln Gly Val Pro Ile Asn Thr Asn Ser Gly
 65 70 75 80
 Pro Asp Asp Gln Ile Gly Tyr Tyr Arg Arg Ala Thr Arg Arg Val Arg
 85 90 95
 Gly Gly Asp Gly Lys Met Lys Glu Leu Ser Pro Arg Trp Tyr Phe Tyr
 100 105 110
 Tyr Leu Gly Thr Gly Pro Glu Ala Ser Leu Pro Tyr Gly Ala Asn Lys
 115 120 125
 Glu Gly Ile Val Trp Val Ala Thr Glu Gly Ala Leu Asn Thr Pro Lys
 130 135 140
 Asp His Ile Gly Thr Arg Asn Pro Asn Asn Asn Ala Ala Thr Val Leu
 145 150 155 160
 Gln Leu Pro Gln Gly Thr Thr Leu Pro Lys Gly Phe Tyr Ala Glu Gly
 165 170 175
 Ser Arg Gly Gly Ser Gln Ala Ser Ser Arg Ser Ser Ser Arg Ser Arg
 180 185 190
 Gly Asn Ser Arg Asn Ser Thr Pro Gly Ser Ser Arg Gly Asn Ser Pro
 195 200 205
 Ala Arg Met Ala Ser Gly Gly Gly Glu Thr Ala Leu Ala Leu Leu Leu
 210 215 220
 Leu Asp Arg Leu Asn Gln Leu Glu Ser Lys Val Ser Gly Lys Gly Gln
 225 230 235 240
 Gln Gln Gln Gly Gln Thr Val Thr Lys Lys Ser Ala Ala Glu Ala Ser
 245 250 255
 Lys Lys Pro Arg Gln Lys Arg Thr Ala Thr Lys Gln Tyr Asn Val Thr
 260 265 270
 Gln Ala Phe Gly Arg Arg Gly Pro Glu Gln Thr Gln Gly Asn Phe Gly
 275 280 285
 Asp Gln Asp Leu Ile Arg Gln Gly Thr Asp Tyr Lys His Trp Pro Gln
 290 295 300
 Ile Ala Gln Phe Ala Pro Ser Ala Ser Ala Phe Phe Gly Met Ser Arg
 305 310 315 320
 Ile Gly Met Glu Val Thr Pro Ser Gly Thr Trp Leu Thr Tyr His Gly
 325 330 335

S226CAS108.ST25

Ala Ile Lys Leu Asp Asp Lys Asp Pro Gln Phe Lys Asp Asn Val Ile
 340 345 350

Leu Leu Asn Lys His Ile Asp Ala Tyr Lys Thr Phe Pro Pro Thr Glu
 355 360 365

Pro Lys Lys Asp Lys Lys Lys Lys Thr Asp Glu Ala Gln Pro Leu Pro
 370 375 380

Gln Arg Gln Lys Lys Gln Pro Thr Val Thr Leu Leu Pro Ala Ala Asp
 385 390 395 400

Met Asp Asp Phe Ser Arg Gln Leu Gln Asn Ser Met Ser Gly Ala Ser
 405 410 415

Ala Asp Ser Thr Gln Ala
 420

<210> 38

<211> 1377

<212> DNA

<213> CORONAVIRUS

<400> 38

atgaagg	tca	ccaaactgct	gcatttagag	acgtacttgt	tgttttaa	aat	aaacgaaca	60
attaaaatgt	ctgataatgg	acccaatca	aaccaacgta	gtgcccccg	cattacattt			120
ggaggacca	cagattcaac	tgacaataac	cagaatggag	gacgcaatgg	ggcaaggcca			180
aaacagcgc	gacccaagg	tttaccat	aatactgcgt	cttggttcac	agctctcact			240
cagcatggca	aggaggaact	tagattccct	cgaggccagg	gcgttccaat	caacaccaat			300
agtgggtccag	atgaccaa	at	tggctactac	cgaagagcta	cccgacgagt	tcgtgggtgg		360
gacggcaaaa	tgaaagagct	cagccccaga	tgggtacttct	attacctagg	aactggcccc			420
gaagcttcac	ttccctacgg	cgctaacaaa	gaaggcatcg	tatgggttgc	aactgagggg			480
gccttgaata	cacccaaga	ccacattggc	accgcaatc	ctaataacaa	tgctgccacc			540
gtgctacaac	ttcctcaagg	aacaacattg	ccaaaaggct	tctacgcaga	gggaagcaga			600
ggcggcagtc	aagcctcttc	tcgctcctca	tcacgtagtc	gcggtaatc	aagaaattca			660
actcctggca	gcagtagggg	aaattctcct	gctcgaatgg	ctagcggagg	tggtgaaact			720
gccctcgcgc	tattgctgct	agacagattg	aaccagcttg	agagcaaagt	ttctggtaaa			780
ggccaacaac	aacaaggcca	aactgtcact	aagaaatctg	ctgctgaggc	atctaaaaag			840
cctcgccaaa	aacgtactgc	cacaaaacag	tacaacgtca	ctcaagcatt	tgaggagacgt			900
ggtccagaac	aaaccaagg	aaatttcggg	gaccaagacc	taatcagaca	aggaactgat			960

S226CAS108.ST25

tacaaacatt	ggccgcaa	at tgcacaattt	gctccaagtg	cctctgcatt	ctttggaatg	1020
tcacgcattg	gcatggaagt	cacaccttcg	ggaacatggc	tgacttatca	tggagccatt	1080
aaattggatg	acaaagatcc	acaattcaaa	gacaacgtca	tactgctgaa	caagcacatt	1140
gacgcataca	aaacattccc	accaacagag	cctaaaaagg	acaaaaagaa	aaagactgat	1200
gaagctcagc	ctttgccgca	gagacaaaag	aagcagccca	ctgtgactct	tcttcctgcg	1260
gctgacatgg	atgattttct	cagacaactt	caaaattcca	tgagtggagc	ttctgctgat	1320
tcaactcagg	cataaacact	catgatgacc	acacaaggca	gatgggctat	gtaaacg	1377

<210> 39

<211> 204

<212> DNA

<213> CORONAVIRUS

<400> 39	atattaggtt	tttacctacc	caggaaaagc	caaccaacct	cgatctcttg	tagatctgtt	60
	ctctaaacga	actttaaaat	ctgtgtagct	gtcgctcggc	tgcatgccta	gtgcacctac	120
	gcagtataaa	caataataaa	ttttactgtc	gttgacaaga	aacgagtaac	tcgtccctct	180
	tctgcagact	gcttacggtt	tcgt				204

<210> 40

<211> 809

<212> DNA

<213> CORONAVIRUS

<400> 40	actcaagcat	ttgggagacg	tggtccagaa	caaacccaag	gaaatttcgg	ggaccaagac	60
	ctaatcagac	aaggaactga	ttacaaacat	tggccgcaaa	ttgcacaatt	tgctccaagt	120
	gcctctgcat	tctttggaat	gtcacgcatt	ggcatggaag	tcacaccttc	gggaacatgg	180
	ctgacttatc	atggagccat	taaattggat	gacaaagatc	cacaattcaa	agacaacgtc	240
	atactgctga	acaagcacat	tgacgcatac	aaaacattcc	caccaacaga	gcctaaaaag	300
	gacaaaaaga	aaaagactga	tgaagctcag	cctttgccgc	agagacaaaa	gaagcagccc	360
	actgtgactc	ttcttcctgc	ggctgacatg	gatgatttct	ccagacaact	tcaaaattcc	420
	atgagtggag	cttctgctga	ttcaactcag	gcataaacac	tcatgatgac	cacacaaggc	480
	agatgggcta	tgtaaacggt	ttcgcaattc	cgtttacgat	acatagtcta	ctcttggtga	540
	gaatgaattc	tcgtaactaa	acagcacaaag	taggtttagt	taactttaat	ctcacatagc	600
	aatctttaat	caatgtgtaa	cattagggag	gacttgaaag	agccaccaca	ttttcatcga	660

S226CAS108.ST25

ggccacgcgg agtacgatcg aggggtacagt gaataatgct agggagagct gcctatatgg	720
aagagcccta atgtgtaaaa ttaatttttag tagtgctatc cccatgtgat tttaatagct	780
tcttaggaga atgacaaaaa aaaaaaaaaa	809

<210> 41

<211> 448

<212> DNA

<213> CORONAVIRUS

<400> 41	
aatgaacaca tagggctgtt caagctgggg cagtacgcct ttttccagct ctactagacc	60
acaagtgcc a tttttgaggt gttcacgtgc ctccgatagg gcctcttcca cagagtcccc	120
gaagccacgc actagcacgt ctctaacctg aaggacaggc aaactgagtt ggacgtgtgt	180
tttctcgttg acaccaagaa caaggctctc catcttacct ttcggtcaca cccggacgaa	240
acctaggtat gctgatgatc gactgcaaca cggacgaaac cgtaagcagt ctgcagaaga	300
gggacgagtt actcgtttct tgtcaacgac agtaaaattt attattgttt atactgcgta	360
ggtgcactag gcatgcagcc gagcgacagc tacacagatt ttaaagttcg tttagagaac	420
agatctacaa gagatcgagg ttggttgg	448

<210> 42

<211> 2033

<212> DNA

<213> CORONAVIRUS

<400> 42	
atacctaggt ttcgtccggg tgtgaccgaa aggtaagatg gagagccttg ttcttggtgt	60
caacgagaaa acacacgtcc aactcagttt gcctgtcctt caggtagag acgtgctagt	120
gcgtggcttc ggggactctg tggaagaggc cctatcggag gcacgtgaac acctcaaaaa	180
tggcacttgt ggtctagtag agctggaaaa aggcgtactg ccccagcttg aacagcccta	240
tgtgttcatt aaacgttctg atgccttaag caccaatcac ggccacaagg tcgttgagct	300
ggttgcagaa atggacggca ttcagtacgg tcgtagcggg ataactctgg gactactcgt	360
gccacatgtg ggcgaaaccc caattgcata ccgcaatgtt cttcttcgta agaacggtaa	420
taagggagcc ggtggtcata gctatggcat cgatctaaag tcttatgact taggtgacga	480
gcttggcact gatccattg aagattatga acaaaactgg aacactaagc atggcagtgg	540
tgcactccgt gaactcactc gtgagctcaa tggagggtgca gtcactcgct atgtcgacaa	600
caatttctgt ggcccagatg ggtaccctct tgattgcata aaagattttc tcgcacgcgc	660

S226CAS108.ST25

```

gggcaagtca atgtgcactc tttccgaaca acttgattac atcgagtcga agagaggtgt 720
ctactgctgc cgtgaccatg agcatgaaat tgcctgggtc actgagcgct ctgataagag 780
ctacgagcac cagacaccct tcgaaattaa gagtgccaaag aaatttgaca ctttcaaagg 840
ggaatgcccc aagtttgtgt ttcctcttaa ctcaaaagtc aaagtcattc aaccacgtgt 900
tgaaaagaaa aagactgagg gtttcatggg gcgtatacgc tctgtgtacc ctgttgcattc 960
tccacaggag tgtaacaata tgcacttgtc taccttgatg aaatgtaatc attgcatga 1020
agtttcatgg cagacgtgcg acttttctgaa agccacttgt gaacattgtg gcactgaaaa 1080
tttagttatt gaaggaccta ctacatgtgg gtacctacct actaatgctg tagtgaaaat 1140
gccatgtcct gcctgtcaag acccagagat tggacctgag catagtgttg cagattatca 1200
caaccactca aacattgaaa ctcgactccg caagggaggt aggactagat gttttggagg 1260
ctgtgtgttt gcctatgttg gctgctataa taagcgtgcc tactgggttc ctcgtgctag 1320
tgctgatatt ggctcaggcc atactggcat tactggtgac aatgtggaga cttgaaatga 1380
ggatctcctt gagatactga gtcgtgaacg tgtaaacatt aacattgttg gcgattttca 1440
tttgaatgaa gaggttgcca tcattttggc atctttctct gcttctacaa gtgcctttat 1500
tgacactata aagagtcttg attacaagtc tttcaaaacc attgttgagt cctgcggtaa 1560
ctataaagtt accaagggaa agcccgtaaa aggtgcttg gacattggac aacagagatc 1620
agttttaaca cactgtgtg gttttccctc acaggctgct ggtgttatca gatcaatttt 1680
tgcgcgcaca cttgatgcag caaaccactc aattcctgat ttgcaaagag cagctgtcac 1740
catacttgat ggtatttctg aacagtcatt acgtcttgtc gacgccatgg tttatacttc 1800
agacctgtc accaacagtg tcattattat ggcataatgta actggtggtc ttgtacaaca 1860
gacttctcag tggttgtcta atcttttggg cactactgtt gaaaaactca ggcctatctt 1920
tgaatggatt gaggcgaaac ttagtgcagg agttgaattt ctcaaggatg cttgggagat 1980
tctcaaattt ctcatcacag gtgtttttga catcgtcaag ggtcaaatac agg 2033

```

<210> 43

<211> 2018

<212> DNA

<213> CORONAVIRUS

<400> 43

```

ggattgaggc gaaacttagt gcaggagttg aatttctcaa ggatgcttgg gagattctca 60
aatttctcat tacaggtgtt tttgacatcg tcaaggggtca aatacagggt gcttcagata 120
acatcaagga ttgtgtaaaa tgcttcattg atgttggtta caaggcactc gaaatgtgca 180
ttgatcaagt cactatcgct ggcgcaaagt tgcgatcact caacttaggt gaagtcttca 240
tcgctcaaag caagggactt taccgtcagt gtatacgtgg caaggagcag ctgcaactac 300

```

S226CAS108.ST25

tcattgcctct	taaggcacca	aaagaagtaa	cctttcttga	aggtgattca	catgacacag	360
tacttacctc	tgaggagggt	gttctcaaga	acggtgaact	cgaagcactc	gagacgcccg	420
ttgatagctt	cacaaatgga	gctatcggtg	gcacaccagt	ctgtgtaaat	ggcctcatgc	480
tcttagagat	taaggacaaa	gaacaatact	gcgcattgtc	tcctggttta	ctggctacaa	540
acaatgtctt	tcgcttaaaa	gggggtgcac	caattaaagg	tgtaaccttt	ggagaagata	600
ctgtttggga	agttcaaggt	tacaagaatg	tgagaatcac	atttgagctt	gatgaacgtg	660
ttgacaaagt	gcttaatgaa	aagtgtctct	tctacactgt	tgaatccggg	accgaagtta	720
ctgagtttgc	atgtgttgta	gcagaggctg	ttgtgaagac	tttacaacca	gtttctgatc	780
tccttaccaa	catgggtatt	gatcttgatg	agtggagtgt	agctacattc	tacttatttg	840
atgatgctgg	tgaagaaaac	ttttcatcac	gtatgtattg	ttccttttac	cctccagatg	900
aggaagaaga	ggacgatgca	gagtgtgagg	aagaagaaat	tgatgaaacc	tgtgaacatg	960
agtacggtac	agaggatgat	tatcaaggtc	tccctctgga	atttggtgcc	tcagctgaaa	1020
cagttcgagt	tgaggaagaa	gaagaggaag	actggctgga	tgatactact	gagcaatcag	1080
agattgagcc	agaaccagaa	cctacacctg	aagaaccagt	taatcagttt	actggttatt	1140
taaaacttac	tgacaatgtt	gccattaaat	gtgttgacat	cgtaaaggag	gcacaaagtg	1200
ctaatectat	gggtgattgta	aatgctgcta	acatacacct	gaaacatggg	gggtggtgtg	1260
caggtgcact	caacaaggca	accaatgggt	ccatgcaaaa	ggagagtgat	gattacatta	1320
agctaaatgg	ccctcttaca	gtaggagggt	cttgtttgct	ttctggacat	aatcttgcta	1380
agaagtgtct	gcatgttggt	ggacctaac	taaatgcagg	tgaggacatc	cagcttctta	1440
aggcagcata	tgaaaatttc	aattcacagg	acatcttact	tgcaccattg	ttgtcagcag	1500
gcatatttgg	tgctaaacca	cttcagtctt	tacaagtgtg	cgtgcagacg	gttcgtacac	1560
aggtttatat	tgcagtcaat	gacaaagctc	tttatgagca	ggttgtcatg	gattatcttg	1620
ataacctgaa	gcctagagtg	gaagcaccta	aacaagagga	gccaccaaac	acagaagatt	1680
ccaaaactga	ggagaaatct	gtcgtacaga	agcctgtcga	tgtgaagcca	aaaattaagg	1740
cctgcattga	tgaggttacc	acaacactgg	aagaaactaa	gtttcttacc	aataagttac	1800
tcttgtttgc	tgatatcaat	ggtaagcttt	accatgattc	tcagaacatg	cttagagggtg	1860
aagatatgtc	tttccttgag	aaggatgcac	cttacatggg	aggtgatgtt	atcactagtg	1920
gtgatatcac	ttgtgttgta	ataccctcca	aaaaggctgg	tggcactact	gagatgctct	1980
caagagcttt	gaagaaagtg	ccagttgatg	agtatata			2018

<210> 44

<211> 1442

<212> DNA

<213> CORONAVIRUS

S226CAS108.ST25

```

<400> 44
ttgatgaggt taccacaaca ctggaagaaa ctaagtttct taccaataag ttactcttgt      60
ttgctgatat caatggtaag ctttaccatg attctcagaa catgcttaga ggtgaagata      120
tgtctttcct tgagaaggat gcaccttaca tggtagggtga tgttatcact agtggtgata      180
tcacttgtgt tgtaataccc tccaaaaagg ctggtggcac tactgagatg ctctcaagag      240
ctttgaagaa agtgccagtt gatgagtata taaccacgta ccctggacaa ggatgtgctg      300
gttatacact tgaggaagct aagactgctc ttaagaaatg caaatctgca ttttatgtac      360
taccttcaga agcacctaag gctaaggaag agattctagg aactgtatcc tggaatttga      420
gagaaatgct tgctcatgct gaagagacaa gaaaattaat gcctatatgc atggatgtta      480
gagccataat ggcaaccatc caacgtaagt ataaaggaat taaaattcaa gagggcatcg      540
ttgactatgg tgtccgattc ttcttttata ctagtaaaga gcctgtagct tctattatta      600
cgaagctgaa ctctctaaat gagccgcttg tcacaatgcc aattgggttat gtgacacatg      660
gttttaaatct tgaagaggct gcgcgctgta tgcgttctct taaagctcct gccgtagtgt      720
cagtatcatc accagatgct gttactacat ataatggata cctcacttcg tcatcaaaga      780
catctgagga gcactttgta gaaacagttt ctttggctgg ctcttacaga gattggtcct      840
attcaggaca gcgtacagag ttaggtgttg aatttcttaa gcgtgggtgac aaaattgtgt      900
accacactct ggagagcccc gtcgagtttc atcttgacgg tgaggttctt tcacttgaca      960
aactaaagag tctcttatcc ctgcgggagg ttaagactat aaaagtgttc acaactgtgg     1020
acaacactaa tctccacaca cagcttgttg atatgtctat gacatatgga cagcagtttg     1080
gtccaacata cttggatggg gctgatgtta caaaaattaa acctcatgta aatcatgagg     1140
gtaagacttt ctttgtacta cctagtgatg acacactacg tagtgaagct ttcgagtact     1200
accatactct tgatgagagt tttcttggtg ggtacatgtc tgcttttaaac cacacaaaga     1260
aatggaaatt tcctcaagtt ggtgggttaa cttcaattaa atgggctgat aacaattgtt     1320
at ttgtctag tgttttatta gcacttcaac agcttgaagt caaattcaat gcaccagcac     1380
ttcaagaggc ttattataga gcccgtgctg gtgatgctgc taacttttgt gcactcatac     1440
tc                                                                    1442

```

<210> 45

<211> 1050

<212> DNA

<213> CORONAVIRUS

```

<400> 45
atatgtctat gacatatgga cagcagtttg gtccaacata cttggatggg gctgatgtta      60
caaaaattaa acctcatgta aatcatgagg gtaagacttt ctttgtacta cctagtgatg      120

```

S226CAS108.ST25

```

acacactacg tagtgaagct ttcgagtact accatactct tgatgagagt tttcttggtta 180
ggtacatgtc tgcttttaaac cacacaaaga aatggaaatt tcctcaagtt ggtgggtttaa 240
cttcaattaa atgggctgat aacaattggtt atttgtctag tgttttatta gcacttcaac 300
agcttgaagt caaattcaat gcaccagcac ttcaagaggc ttattataga gcccgtgctg 360
gtgatgctgc taacttttgt gcactcatac tgccttacag taataaaaact gttggcgagc 420
ttggtgatgt cagagaaaact atgacccatc ttctacagca tgctaatttg gaatctgcaa 480
agcgagttct taatgtggtg tgtaaacatt gtggtcagaa aactactacc ttaacggggtg 540
tagaagctgt gatgtatatg ggtactctat cttatgataa tcttaagaca ggtgtttcca 600
ttccatgtgt gtgtggctgt gatgctacac aatatctagt acaacaagag tcttcttttg 660
ttatgatgtc tgcaccacct gctgagtata aattacagca aggtacattc ttatgtgcga 720
atgagtacac tggtaactat cagtgtggtc attacactca tataactgct aaggagaccc 780
tctatcgtat tgacggagct caccttaca agatgtcaga gtacaaagga ccagtgactg 840
atgttttcta caaggaaaca tcttacctta caaccatcaa gcctgtgtcg tataaaactcg 900
atggagttac ttacacagag attgaaccaa aattggatgg gtattataaa aaggataatg 960
cttactatac agagcagcct atagaccttg taccaactca accattacca aatgcgagtt 1020
ttgataattt caaactcaca tgttctaaca 1050

```

<210> 46

<211> 1995

<212> DNA

<213> CORONAVIRUS

<400> 46

```

tttgtgcact catactcgct tacagtaata aaactgttgg cgagcttggt gatgtcagag 60
aaactatgac ccatcttcta cagcatgcta atttgaatc tgcaaagcga gttcttaatg 120
tggtgtgtaa acattgtggt cagaaaacta ctaccttaac ggggtgtagaa gctgtgatgt 180
atatgggtac tctatcttat gataatctta agacaggtgt ttccattcca tgtgtgtgtg 240
gtcgtgatgc tacacaatat ctagtacaac aagagtcctc ttttgttatg atgtctgcac 300
cacctgctga gtataaatta cagcaaggta cattcttatg tgcgaaatgag tacactggta 360
actatcagtg tggtcattac actcatataa ctgctaagga gaccctctat cgtattgacg 420
gagctcacct tacaagatg tcagagtaca aaggaccagt gactgatgtt ttctacaagg 480
aaacatctta cactacaacc atcaagcctg tgtcgtataa actcgatgga gttacttaca 540
cagagattga accaaaattg gatgggtatt ataaaaagga taatgcttac tatacagagc 600
agcctataga cttgtacca actcaaccat taccaaatgc gagttttgat aatttcaaac 660
tcacatgttc taacacaaaa tttgctgatg atttaaataa aatgacaggc ttcacaaagc 720

```

S226CAS108.ST25

```

cagcttcacg agagctatct gtcacattct tcccagactt gaatggcgat gtagtggcta 780
ttgactatag acactattca gcgagtttca agaaagggtgc taaattactg cataagccaa 840
ttgtttggca cattaaccag gctacaacca agacaacggt caaaccaaac acttggtggt 900
tacgttgtct ttggagtaca aagccagtag atacttcaaa ttcatttgaa gttctggcag 960
tagaagacac acaaggaatg gacaatcttg cttgtgaaag tcaacaaccc acctctgaag 1020
aagtagtgga aaatcctacc atacagaagg aagtcataga gtgtgacgtg aaaactaccg 1080
aagttgtagg caatgtcata cttaaaccat cagatgaagg tggttaaagta acacaagagt 1140
taggtcatga ggatcttatg gctgcttatg tgaaaaacac aagcattacc attaagaaac 1200
ctaattgagct ttcactagcc ttaggtttaa aaacaattgc cactcatggt attgctgcaa 1260
ttaatagtgt tccttggagt aaaattttgg cttatgtcaa accattctta ggacaagcag 1320
caattacaac atcaaattgc gctaagagat tagcacaacg tgtgtttaac aattatatgc 1380
cttatgtggt tacattattg ttccaattgt gtacttttac taaaagtacc aattctagaa 1440
ttagagcttc actacctaca actattgcta aaaatagtgt taagagtgtt gctaaattat 1500
gtttggatgc cggcattaat tatgtgaagt cacccaaatt ttctaaattg ttcacaatcg 1560
ctatgtggct attgtttgta agtatttgct taggttctct aatctgtgta actgctgctt 1620
ttggtgtact cttatctaatt tttggtgctc cttcttattg taatggcggt agagaattgt 1680
atcttaattc gtctaacggt actactatgg atttctgtga aggttctttt ccttgagca 1740
tttgtttaag tggattagac tcccttgatt cttatccagc tcttgaaacc attcaggtga 1800
cgatttcacg gtacaagcta gacttgacaa ttttaggtct ggccgctgag tgggttttgg 1860
catatatgtt gttcacaaaa ttcttttatt tattaggtct ttcagctata atgcaggtgt 1920
tctttggcta ttttgctagt catttcatca gcaattcttg gctcatgtgg tttatcatta 1980
gtattgtaca aatgg 1995

```

<210> 47

<211> 1884

<212> DNA

<213> CORONAVIRUS

<400> 47

```

aattcttggc tcatgtgggt tatcattagt attgtacaaa tggcacccgt ttctgcaatg 60
gttaggatgt acatcttctt tgcttctttc tactacatat ggaagagcta tgttcatatc 120
atggatgggt gcacctcttc gacttgcatg atgtgctata agcgcaatcg tgccacacgc 180
gttgagtgtg caactattgt taatggcatg aagagatctt tctatgtcta tgcaaatgga 240
ggccgtggct tctgcaagac tcacaattgg aattgtctca attgtgacac attttgact 300
ggtagtacat tcattagtga tgaagttgct cgtgatttgt cactccagtt taaaagacca 360

```

S226CAS108.ST25

```

atcaacccta ctgaccagtc atcgatatatt gttgatagtg ttgctgtgaa aaatggcgcg 420
cttcacctct actttgacaa ggctgggtcaa aagacctatg agagacatcc gctctcccat 480
tttgtcaatt tagacaattt gagagctaac aacactaaag gttcactgcc tattaatgtc 540
atagtttttg atggcaagtc caaatgcgac gagtctgctt ctaagtctgc ttctgtgtac 600
tacagtcagc tgatgtgccca acctattctg ttgcttgacc aagctcttgt atcagacggt 660
ggagatagta ctgaagtttc cgtaagatg tttgatgctt atgtcgacac cttttcagca 720
acttttagtg ttcctatgga aaaacttaag gcacttggtg ctacagctca cagcgagtta 780
gcaaaggggtg tagcttttaga tgggtgtcctt tctacattcg tgtcagctgc cgcacaaggt 840
gttggttgata ccgatgttga cacaaaggat gttattgaat gtctcaaact ttcacatcac 900
tctgacttag aagtgcagag tgacagttgt aacaatttca tgctcaccta taataagggt 960
gaaaacatga cgcccagaga tcttggcgca tgtattgact gtaatgcaag gcatatcaat 1020
gcccagtag caaaaagtca caatgtttca ctcactctgga atgtaaaaga ctacatgtct 1080
ttatctgaac agctgcgtaa acaaattcgt agtgctgcca agaagaacaa catacctttt 1140
agactaactt gtgctacaac tagacagggt gtcaatgtca taactactaa aatctcactc 1200
aagggtggta agattgttag tacttgtttt aaacttatgc ttaaggccac attattgtgc 1260
gttcttgctg cattggtttg ttatatcggt atgccagtac atacattgtc aatccatgat 1320
ggttacacaa atgaaatcat tggttacaaa gccattcagg atggtgtcac tcgtgacatc 1380
atctctactg atgattgttt tgcaaataaa catgctgggt ttgacgcatg gtttagccag 1440
cgtgggtggtt catacaaaaa tgacaaaagc tgccctgtag tagctgctat cattacaaga 1500
gagattgggt tcatagtgcc tggcttaccg ggtactgtgc tgagagcaat caatgggtgac 1560
ttcttgcat tttctacctg tgtttttagt gctgttgcca acatttgcta cacaccttcc 1620
aaactcattg agtatagtga ttttgctacc tctgcttgcg ttcttgctgc tgagtgtaca 1680
atttttaagg atgctatggg caaacctgtg ccatattggt atgacactaa tttgctagag 1740
ggttctatct cttatagtga gcttcgtcca gacactcggt atgtgcttat ggatgggtcc 1800
atcatacagt ttcctaacac ttacctggag gggtctgtta gagtagtaac aacttttgat 1860
gctgagtact gtagacatgg taca 1884

```

<210> 48

<211> 2020

<212> DNA

<213> CORONAVIRUS

<400> 48

```

cactcggtat gtgcttatgg atggttccat catacagttt cctaacactt acctggaggg 60
ttctgttaga gtagtaacaa cttttgatgc tgagtactgt agacatggta catgcgaaag 120

```

S226CAS108.ST25

```

gtcagaagta ggtatttgcc tatctaccag tggtagatgg gttcttaata atgagcatta 180
cagagctcta tcaggagttt tctgtggtgt tgatgcatg aatctcatag ctaacatctt 240
tactcctctt gtgcaacctg tgggtgcttt agatgtgtct gcttcagtag tggctggtgg 300
tattattgcc atattggtga cttgtgctgc ctactacttt atgaaattca gacgtgtttt 360
tggtgagtac aaccatgttg ttgctgctaa tgcacttttg tttttgatgt ctttcactat 420
actctgtctg gtaccagctt acagctttct gccgggagtc tactcagtct tttacttgta 480
cttgacattc tttttacca atgatgtttc attcttggct caccttcaat ggtttgccat 540
gttttctcct attgtgcctt tttggataac agcaatctat gtattctgta tttctctgaa 600
gcactgccat tggttcttta acaactatct taggaaaaga gtcattgtta atggagttac 660
atttagtacc ttcgaggagg ctgctttgtg tacctttttg ctcaacaagg aaatgtacct 720
aaaattgcgt agcgagacac tgttgccact tacacagtat aacaggatc ttgctctata 780
taacaagtac aagtatttca gtggagcctt agatactacc agctatcgtg aagcagcttg 840
ctgccactta gcaaaggctc taaatgactt tagcaactca ggtgctgatg ttctctacca 900
accaccacag acatcaatca cttctgctgt tctgcagagt ggtttttaga aaatggcatt 960
cccgtcaggc aaagtgaag ggtgcatggt acaagtaacc tgtggaacta caactcttaa 1020
tggattgtgg ttggatgaca cagtatactg tccaagacat gtcatttgca cagcagaaga 1080
catgcttaat cctaactatg aagatctgct cattcgcaa tccaaccata gctttcttgt 1140
tcaggctggc aatgttcaac ttcgtgttat tggccattct atgcaaaatt gtctgcttag 1200
gcttaaagtt gatacttcta accctaagac acccaagtat aaatttgtcc gtatccaacc 1260
tgggtcaaca ttttcagttc tagcatgcta caatggttca ccatctggtg tttatcagtg 1320
tgccatgaga cctaatacata ccattaaagg ttctttcctt aatggatcat gtggtagtgt 1380
tggttttaac attgattatg attgctgtgc tttctgctat atgcatcata tggagcttcc 1440
aacaggagta cacgctggta ctgacttaga aggtaaattc tatggtccat ttgttgacag 1500
acaaactgca caggctgcag gtacagacac aaccataaca ttaaatgttt tggcatggct 1560
gtatgctgct gttatcaatg gtgatagggt gtttcttaat agattcacca ctactttgaa 1620
tgactttaac cttgtggcaa tgaagtacaa ctatgaacct ttgacacaag atcatgttga 1680
catattggga cctctttctg ctcaaacagg aattgccgtc ttagatatgt gtgctgcttt 1740
gaaagagctg ctgcagaatg gtatgaatgg tcgtactatc cttggtagca ctattttaga 1800
agatgagttt acaccatttg atgttgtag acaatgctct ggtgttacct tccaaggtaa 1860
gttcaagaaa attgttaagg gcactcatca ttggatgctt ttaactttct tgacatcact 1920
attgattctt gttcaagta cacagtggct actgtttttc tttgtttacg agaatgcttt 1980
cttgccattt actcttggtt ttatggcaat tgctgcatgt 2020

```

<210> 49

S226CAS108.ST25

<211> 2040

<212> DNA

<213> CORONAVIRUS

<400> 49
agcatttcca gcctgaagac gtactgtagc agctaaactg cccagcacca tacctctatt 60
taggttgttt aagcctttga tgaagtacaa gtatttact ttaggccctt ttggtgtgtc 120
tgtaacaaac ctacaagggtg gttccagttc tgtgtaaatt gtacctgtac catcactctt 180
agggaatcta gcccatttga gatcttggtg gtctgatagt aatgccagca caaacctacc 240
tcccttcgaa ttgttatagt aggcaagtgc attgtcatca gtacaagctg tttgtgtggt 300
accagccgca caggacatct gtcgtagtgc tactggactc agttcattat tctgtagttt 360
aacagctgag ttggctctta gagctgtaac aataagaggc caagccaaat ttggtgaatt 420
gtccatgtta atttacttaa gttgaacaat cttgctatcc gcatcaacaa cttgctggat 480
ttcccagagt gcagatgcat atgtaaagggt gttaccatca caagtgttct tgtaggtacc 540
ataatcaggg acaacaacca tgagtttggc tgctgtagtc aatggatga tggtgagtgg 600
aacacaacca tcacgcgcat tgttgataat gttgttaagt gcatcattat caagcttcct 660
aagcatagtg aagagcattg tttgcatagc actagttact tttgccctct tgcctcaga 720
tcttgccctgt ttgtacattt gggcatagc ctgatctgcc atcttttcca acttgcgttg 780
catggcagca tcacgggtcaa actcagattt agccacattc aaagatttct ttaacttttt 840
gagaacgact tcagaatcac cattagctac agcctgctca taggcctcct gggcagtggc 900
ataagcggca tatgatggta aagaactaaa ttctgaagca atagcctgaa gagtagcacg 960
gttatcgagc atttcctcgc acaacctatt aatgtctaca gcaccctgca tggatagcaa 1020
aacagacaaa agagaaacca tcttctcgaa agcttcagtt gtgtcttttg caagaagaat 1080
atcattgttg agttgtacac attgtgcccc caatttagaa gatgactcta ctctaagttg 1140
ttgaagaacc gagagcagta ccacagatgt gcactttacg tcagacattt tagactgtac 1200
agtagcaacc ttgatacatg gtttacctcc aatacccaac aacttaatgt taagcttgaa 1260
agcatcaata ctactcttag gaggcaaaag cccctgggag ttcatatacc taaattcttg 1320
tgtagagacc aagtagtcat aaacaccaag agtaagcctg aagtaacggt tgagtaaaca 1380
gaaaaggcca aagtagcagc agcaacaata gcctaagaaa caataaaca gcatgatata 1440
ctgtaagggtg ttgccagtaa taaataaca tgggtaatac tcaacacaca caaacactat 1500
agctctagct aaaaacatga tagtcgtaac gacaccagaa tagttagagg ttacagaaat 1560
aactaaggcc cacatggaaa tagcttgatc taaagcatta ccatagtaga ctttgtaaac 1620
aagtgtaatg acattcatca gtgtccaaac acgtctagca gcatcatcat aaacagtgcg 1680
agctgtcatg agaataagca aaactaaagc tgaagcatac ataacacaat ccttaagcct 1740
ataaccagac aagctagtgt cagccaattc aagccatgtc atgatacgca tcaccagct 1800

S226CAS108.ST25

agcaggcatg	tagaccatat	taaagtaagc	aactgtttgca	agagaaggta	acagaaacaa	1860
gcacaagaat	gcgtgcttat	gcttaacaag	cagcatagca	catgcagcaa	ttgccataat	1920
accaagagta	aatggcaaga	aagcattctc	gtaaacaaag	aaaaacagtg	accactgtgt	1980
actttgaaca	agaatcaata	gtgatgtcaa	gaaagttaaa	agcatccaat	gatgagtgca	2040

<210> 50

<211> 2012

<212> DNA

<213> CORONAVIRUS

<400> 50

cttgtagggt	tggtacagac	acaccaaaaag	ggcctaaagt	gaaataacttg	tactttcatca	60
aaggcttaaa	caacctaaat	agaggatatg	tgctgggcag	tttagctgct	acagtacgtc	120
ttcaggctgg	aaatgctaca	gaagtacctg	ccaattcaac	tgtgctttcc	ttctgtgctt	180
ttgcagtaga	ccctgctaaa	gcatataagg	attacctagc	aagtggagga	caaccaatca	240
ccaactgtgt	gaagatgttg	tgtacacaca	ctggtacagg	acaggcaatt	actgtaacac	300
cagaagctaa	catggaccaa	gagtcctttg	gtggtgcttc	atgttgtctg	tattgtagat	360
gccacattga	ccatccaaat	cctaaaggat	tctgtgactt	gaaaggtaag	tacgtccaaa	420
tacctaccac	ttgtgctaata	gaccagtg	gttttacact	tagaaacaca	gtctgtaccg	480
tctgcggaat	gtggaaagg	tatggctgta	gttgtagacca	actccgcgaa	cccttgatgc	540
agtctgcgga	tgcatcaacg	tttttaaacg	ggtttgcggt	gtaagtgcag	cccgtcttac	600
accgtgcggc	acaggcacta	gtactgatgt	cgtctacagg	gcttttgata	tttacaacga	660
aaaagttgct	ggttttgcaa	agttcctaaa	aactaattgc	tgctcgcttcc	aggagaagga	720
tgaggaaggc	aattttattag	actcttactt	tgtagttaag	aggcatacta	tgtctaacta	780
ccaacatgaa	gagactat	ataacttgg	taaagattgt	ccagcggttg	ctgtccatga	840
ctttttcaag	tttagagtag	atggtgacat	ggtaccacat	atatcacgtc	agcgtctaac	900
taaatacaca	atggctgatt	tagtctatgc	tctacgtcat	tttgatgagg	gtaattgtga	960
tacattaaaa	gaaataactg	tcacatacaa	ttgctgtgat	gatgattatt	tcaataagaa	1020
ggattgggtat	gacttcgtag	agaatcctga	catcttacgc	gtatatgcta	acttaggtga	1080
gcgtgtacgc	caatcattat	taaagactgt	acaattctgc	gatgctatgc	gtgatgcagg	1140
cattgtaggc	gtactgacat	tagataatca	ggatcttaat	gggaactggt	acgatttcgg	1200
tgatttcgta	caagtagcac	caggctgcgg	agttcctatt	gtggattcat	attactcatt	1260
gctgatgcc	atcctcactt	tgactagggc	attggctgct	gagtcccata	tggatgctga	1320
tctcgcaaaa	ccacttatta	agtgggattt	gctgaaatat	gattttacgg	aagagagact	1380
ttgtctcttc	gaccgttatt	ttaaatattg	ggaccagaca	taccatccca	attgtattaa	1440

S226CAS108.ST25

```

ctgtttggat gataggtgta tccttcattg tgcaaaacttt aatgtgttat tttctactgt 1500
gtttccacct acaagttttg gaccactagt aagaaaaata tttgtagatg gtgttccttt 1560
tgttgtttca actggatacc attttcgtga gttaggagtc gtacataatc aggatgtaaa 1620
cttacatagc tcgcgtctca gtttcaagga acttttagtg tatgctgctg atccagctat 1680
gcatgcagct tctggcaatt tattgctaga taaacgcact acatgctttt cagtagctgc 1740
actaacaac aatgttgctt ttcaaactgt caaacccggg aattttaata aagactttta 1800
tgactttgct gtgtctaaag gtttctttaa ggaaggaagt tctgttgaac taaaacactt 1860
cttctttgct caggatggca acgctgctat cagtgattat gactattatc gttataatct 1920
gccaacaatg tgtgatatca gacaactcct attcgtagtt gaagttgttg ataaataactt 1980
tgattgttac gatggtggct gtattaatgc ca 2012

```

<210> 51

<211> 1877

<212> DNA

<213> CORONAVIRUS

```

<400> 51
gtacttcgcg tacagtggca ataccatatg acagcttaaa tgtttcctca gtggctttga 60
gcgtttctgc tgcgaaaagc ttgagtctct cagtacaagt gtgggcaagt atgtaatcgc 120
cagcattagt ccaatcacat gttgctatcg cattgaagtc agtgacattg tctgccta 180
cacatgtgtt tttgtataaa ccaaaaacct gaccattagc acataatgga aaactaatgg 240
gaggcttatg tgacttgcaa taatagctca tacctcctag atacagttgt gtcacatcag 300
tgacatcaca acctgggggca ttgcaaacat agggattaac agacaacact aatttgtgtg 360
atgttgaaat gacatggtca tagcagcact tgcaacatag gaatggtctc ctaatacagg 420
caccgcaacg aagtgaagtc tgtgaattgc acaatacaca agcacctaca gcctgcaaga 480
ctgtatgtgg tgtgtacata gcctcataaa actcagggtc ccagtaccgt gaggtgttat 540
cattagttag cattacggaa tacatgtcca acatgtggcc agtaagctca tcatgtaact 600
ttctaatagta ttgtaaatac aagtgaagaa catcagcata ctcctgatta ggatgttttg 660
taagtgggta agcatcaata gccagtgaca cgaacctttc aatcataagt gtaccatctg 720
ttttgacaat atcatcgaca aaacagcctg cgctaatat tcttgatgga tctgggtaag 780
gcagggtacac gtaatcatct cttgttttaa ctagcattgt atgctgtgag caaaattcgt 840
gaggtccttt agtaagggtca gtctcagtc aacattttgc ctcagacatg aacacattat 900
tttgataata aagaactgcc ttaaagttct taatgctagc tactaaacct tgagccgcat 960
agttactgtt atagcacaca acggcatcat cagaaagaat catcatggag aaatgtttac 1020
gcaggtaagc gtaaaactca tccacgaatt catgatcaac atccctatct ctatagagac 1080

```

S226CAS108.ST25

actcatagag cctgtgttgt agattgcgga catacttgtc agctatctta ttaccatcag	1140
ttgaaagaag tgcatttaca ttggctgtaa cagcttgaca aatgttaaag acactattag	1200
cataagcagt tgtagcatca ccggatgatg ttccacctgg tttaacatat agtgagccgc	1260
cacacatgac catctcactt aatacttgcg cacactcggt agctaacctg tagaaacggt	1320
gtgataagtt acagcaagtg ttatgtttgc gagcaagaac aagagaggcc attatcctaa	1380
gcatgttagg catggctctg tcacattttg gataatccca acccataagg tgtggagttt	1440
ctacatcact gtaaacagtt tttaacatat tatgccagcc accgtaaac ttgcttggtc	1500
caattaccac agtagctcct ctagtggcgg ctattgactt caataatttc tgatgaaact	1560
gtctatttgt catagtacta cagatagaga caccagctac ggtgcgagct ctattctttg	1620
cactaatggc atacttaaga ttcatttgag ttatagtagg gatgacatta cgcttagtat	1680
acgcgaaaag tgcattctga tcctcataac tcattgagtc ataataaagt ctagccttac	1740
cccatttatt aaatgggaaa ccagctgatt tatccagatt gttaacgatt acttggttgg	1800
cattaataca gccaccatcg taacaatcaa agtatattatc aacaacttca actacgaata	1860
ggagttgtct gatatca	1877

<210> 52

<211> 2051

<212> DNA

<213> CORONAVIRUS

<400> 52

tcaggtccaa tcttgacaaa gtacttcatt gatgtaagct caaagccatg cgcccaaagg	60
acgaacacga ctctgtctga caatcctttc agtgtatcac tgagcatttg tactatctta	120
atacgacta cattccaggg caagccttta tacatgagtg gtataagatg tttaaactgg	180
tcacctggtg gaggttttgc attaactctg gtgaattctg tgttattttc agtgtcaaca	240
taaccagtcg gtacagctac taagttaaca cctgtagaaa atcctagctg gagaggtagg	300
ttagtaccga cagcatctct agttgcatga cagccctcta catcaaagcc aatccacgca	360
cgaacgtgac gaatagcttc ttcgcgggtg ataaacatat tagggtaacc attgacttgg	420
taattcattt tgaaacccat catagagatg agtctacggt aggtcatgtc ctttggtatg	480
cctggatgtg caacacataa tccttcagtc ttgaacttta tatcaacgct gaggtgtgta	540
ggtgcctgtg taggatgaag accagtaatg atcttactac agtccttaaa aagtccagtt	600
acattttctg cttgtaatgt agccacattg cgacgtggta tttctagact tgtaaattgc	660
agtttgtcat aaagatctct atcagacatt atgcacaaaa tgccaatttt tgcccttgtg	720
atagccacat tgaagcgggt gacattacaa gagtgtgctg tttcagtagt ttgtgtgaat	780
atgacatagt catattcaga accctgtgat gaatcaacag tctgcgtagg caatcctaag	840

S226CAS108.ST25

```

atTTTTgaag ctacagcgTT ctgtgaatta taaggtgaga taaaaacagc ttttctccaa 900
gcaggattgc gtgtaagaaa ttctcttaca acgcctatTT gaggtctgtt gattgcagat 960
gaaacatcat gtgtaataac acctttgtag aacatTTTga agcattgagc tgacttatcc 1020
ttgtgtgctt ttagcttatt gtcataaaact aaagcactca cagtgtcaac aatttcagca 1080
ggacaacggc gacaagttcc aaggaacatg tctggacctt ttgttttcat aagtctgcac 1140
actgaattaa aatattcttg ttctagtgtg cctttagtca gcaatgtgcg gggggctggt 1200
aattgagcag gatcgccaat atagacgtag tgTTTTgcac gaagtctagc attgacaaca 1260
ctcaagtcac aattagtagc catagagatt tcatcaaaga ctacaatgtc agcagttgtt 1320
tctggcaatg catttacagt gcagaaaaca tactgttcta gtgttgaatt cactttgaat 1380
ttatcaaaac actctacgcg cgcacgcgca ggtatgattc tactacatTT atctatgggc 1440
aaatattTTa atgcctTTTc acatagggca tcaacagctg catgagagca tgccgtatac 1500
actatgcbag cagatgggta atagagagca agtccgatgg caaaatgact cttaccagta 1560
ccaggTggTc cttggagtgt agagtactTT tgcatgccga cttttgata atttgcaaca 1620
ttgctagaaa actcatctga gatgttgagt gttgggtaca agccagtaat tctcacatag 1680
tgctcttTgt gcactagagt aggtgcacta agtggcatta cagtgtgaga tgtcaacaca 1740
aagtaatcac caacattcaa cttgtatgtc gtagtacctc tgtacacaac agcatcacca 1800
tagtcacctt tttcaaaggT gtactctcca atctgtactt tactatTTTt agttacacgg 1860
taaccagtaa agacatagtt tctgttcaat ggtggTctag gttttccaac ctcccatgaa 1920
agatgcaatt ctctgtcaga gagtacttcg cgtacagtgg caataccata tgacagctta 1980
aatgtttcct cagtggctTT gagcgtttct gctgcgaaaa gcttgagtct ctcagtacaa 2040
gtgTtggcaa g 2051

```

<210> 53

<211> 2075

<212> DNA

<213> CORONAVIRUS

<400> 53

```

tgcttTtagt tttgggtaga aggtttcaac atgtccatcc ttacaccaaa gcatgaatga 60
aatttcagca tagtcaattg taaccttgac cacttttgaa atcactgaca aatcttTtga 120
ctttattatc tcgacaaagt catcaagtaa aagatcaatc acagaacaca cacattttga 180
tgaacctgtt tgcbcatctg ttatgaagta atttttcact gtgctgtcca tagggataaa 240
atcctctaatt ttaagtggTg aatcttTtga gcgcttggct aagcctatca ttaaTgaag 300
accgccaagt tgtccatgac tgaaatctcc ataaacgatg tgTtcgaagg catagccctc 360
gagcttatat cgctgtatga attcatccat agcgagctcg agaaagtcaT tttccatttg 420

```

S226CAS108.ST25

tgatctgggc	ttaaaatcct	ctaagtctct	gctctgagta	aagtaggttt	caggcaactg	480
ttgaataatg	ccgtctactt	tcttaaagta	gttaaactgt	gtttttactg	attctccaat	540
taatgtgact	ccattgacgc	tagcttgtgc	tgggtccctt	gaagggtgta	gacctttgac	600
tgaaccttct	gttattaaaa	caccattacg	ggcgtttcta	aaaagggtcta	cctgtccttc	660
cactctacca	tcaaacaaga	cagtaagtga	agaacaagca	ctctcagtag	gtttcttggc	720
aatgtcagtc	attgtgcaga	cacctattgt	agatacatgt	gctggggctt	ctcttttgta	780
gtcccagatt	acagtattag	cagcgatatc	aacacccaaa	ttattgagta	tcttaatctc	840
tggcactggg	ttaatgttac	gcttagccca	aagctcaa	gcaacattaa	caggaagtg	900
tgtcttattt	tcaaagatct	ccacatcaat	accatctacc	tttgtgtaaa	cagcattatt	960
aatgatggaa	acagggtgctt	cgccggcggtg	tccatcaaag	tgtcctttat	taacaacatt	1020
ataagccaca	ttttctaaac	tctgtaacct	ggtaaatgta	ttccacaggt	tataagtatc	1080
aaattgtttg	taaatccata	ggctaaatcc	agcagaaatc	atcatattat	atgcatccaa	1140
gtactgtcgg	tactcatttg	catgggtgtct	gcaaacagca	ccacctaaat	tgcacgtgtg	1200
aatacacgta	gcagatttga	gtggaacata	atcaatatcc	gacactactt	gtttgccatg	1260
agactcacia	ggactatcag	aatagtaaaa	gaaaggcaat	tgttttaaat	tagtaaatgc	1320
acttttatcg	aaagctggag	tgtggaatgc	atgcttattc	acatacaaac	taccaccatc	1380
acagcctggg	aagttcaagt	ttgacaagac	tcttgtgtca	aacctacaca	caattgcatt	1440
ggctgggtaa	cgatcaacgt	tacaattcca	aaacaaacaa	acaccatcag	tgaatttatc	1500
gtgatgtgta	gcataagaat	agaagagttc	ctctattttg	taagctttgt	cactacatgg	1560
ctgagcatcg	tagaacttcc	attctacttc	agcctgaggc	acacacttga	tagcctttgg	1620
atttccaatg	tcatgaagaa	ctggaaactt	atcagcaagc	aatgcagact	tcacaacat	1680
gtgttgact	tttctgcaag	cagaattaac	cctcagttca	tctcctataa	taggggtattc	1740
aacagaccaa	tcaacgcgct	taacaaagca	ctcatggact	gctaaacatc	tagtcatgat	1800
agcatcacia	ctagccacat	gtgcatttcc	atgtacctgg	caatgttggt	catgggtact	1860
ctgaagggtta	cccgtaaagc	cccactgctg	aacatcaatc	ataaatgggt	tatagacata	1920
gtcaaaaccc	acagaatgat	tccagcaggc	ataagtatct	gatgaagtag	aaaagcaagt	1980
tgcacgtttg	tcacacagac	aacacgttct	ttcaggtcca	atcttgacaa	agtacttcat	2040
tgatgtaagc	tcaaagccat	gcgcccacaa	gacga			2075

<210> 54

<211> 1891

<212> DNA

<213> CORONAVIRUS

S226CAS108.ST25

<400> 54
 aagattcacc acttaaatta gaggatttta tccctatgga cagcacagtg aaaaattact 60
 tcataacaga tgcgcaaaca ggttcatcaa aatgtgtgtg ttctgtgatt gatcttttac 120
 ttgatgactt tgtcgagata ataaagtcac aagatttgtc agtgatttca aaagtgggtca 180
 aggttacaat tgactatgct gaaatttcat tcatgctttg gtgtaaggat ggacatgttg 240
 aaaccttcta cccaaaacta caagcaagtc aagcgtggca accagggtgtt gcgatgccta 300
 acttgtacaa gatgcaaaga atgcttcttg aaaagtgtga ccttcagaat tatggtgaaa 360
 atgctgttat accaaaagga ataatgatga atgtcgcaaa gtataactcaa ctgtgtcaat 420
 acttaaatac acttacttta gctgtaccct acaacatgag agttattcac tttggtgctg 480
 gctctgataa aggagttgca ccagggtacag ctgtgctcag acaatgggtg ccaactggca 540
 cactacttgt cgattcagat cttaatgact tcgtctccga cgcagattct actttaattg 600
 gagactgtgc aacagtacat acggctaata aatgggacct tattattagc gatatgtatg 660
 accctaggac caaacatgtg acaaaagaga atgactctaa agaagggttt ttcacttatc 720
 tgtgtggatt tataaagcaa aaactagccc tgggtgggtc tatagctgta aagataacag 780
 agcattcttg gaatgctgac ctttacaagc ttatgggcca tttctcatgg tggacagctt 840
 ttgttacaaa tgtaaatgca tcatcatcgg aagcattttt aattggggct aactatcttg 900
 gcaagccgaa ggaacaaatt gatggctata ccatgcatgc taactacatt ttctggagga 960
 acacaaatcc tatccagttg tcttcctatt cactctttga catgagcaaa tttcctctta 1020
 aattaagagg aactgctgta atgtctctta aggagaatca aatcaatgat atgatttatt 1080
 ctcttctgga aaaaggtagg cttatcatta gagaaaacaa cagagttgtg gtttcaagtg 1140
 atattcttgt taacaactaa acgaacatgt ttattttctt attatttctt actctcacta 1200
 gtggtagtga ccttgaccgg tgcaccactt ttgatgatgt tcaagctcct aattacactc 1260
 aacatacttc atctatgagg ggggtttact atcctgatga aattttttaga tcagacactc 1320
 tttatttaac tcaggattta tttcttccat tttattctaa tgttacaggg tttcatacta 1380
 ttaatcatac gtttggcaac cctgtcatac cttttaagga tgggtatttat tttgctgcca 1440
 cagagaaatc aaatgttgtc cgtggttggg tttttggttc taccatgaac aacaagtcac 1500
 agtcggtgat tattattaac aatttacta atgttgttat acgagcatgt aactttgaat 1560
 tgtgtgacaa ccctttcttt gctgtttcta aacccatggg tacacagaca catactatga 1620
 tattcgataa tgcatttaat tgcactttcg agtacatatc tgatgccttt tcgcttgatg 1680
 tttcagaaaa gtcaggtaat tttaaacact tacgagagtt tgtgtttaaa aataaagatg 1740
 ggtttctcta tgtttataag ggctatcaac ctatagatgt agttcgtgat ctaccttctg 1800
 gttttaacac tttgaaacct atttttaagt tgcctcttgg tattaacatt acaaatttta 1860
 gagccattct tacagccttt tcacctgctc a 1891

<210> 55

S226CAS108.ST25

<211> 32

<212> DNA

<213> ARTIFICIAL SEQUENCE

<220>

<223> amorce N sens

<400> 55

cccatatgtc tgataatgga cccaatcaa ac

32

<210> 56

<211> 32

<212> DNA

<213> artificial sequence

<220>

<223> amorce N antisens

<400> 56

cccccggtg cctgagttga atcagcagaa gc

32

<210> 57

<211> 31

<212> DNA

<213> artificial sequence

<220>

<223> amorce Sc sens

<400> 57

cccatatgag tgaccttgac cggcgcacca c

31

<210> 58

<211> 30

<212> DNA

<213> artificial sequence

<220>

<223> amorce SL sens

<400> 58

cccatatgaa accttgacc ccacctgctc

30

S226CAS108.ST25

<210> 59

<211> 33

<212> DNA

<213> amorce Sc et SL antisens

<400> 59

cccccggtt taatatattg ctcataattt ccc

33

<210> 60

<211> 16

<212> DNA

<213> amorce sens série 1

<400> 60

ggcatcgtat gggttg

16

<210> 61

<211> 16

<212> DNA

<213> amorce antisens série 2 (28774-28759)

<400> 61

cagtttcacc acctcc

16

<210> 62

<211> 16

<212> DNA

<213> amorce sens série 2 (28375-28390)

<400> 62

ggctactacc gaagag

16

<210> 63

<211> 16

<212> DNA

<213> amorce antisens série 2 (28702-28687)

S226CAS108.ST25

<400> 63
aattaccgcg actacg 16

<210> 64

<211> 26

<212> DNA

<213> sonde 1/série 1 (28561-28586)

<400> 64
ggcaccgcga atcctaataa caatgc 26

<210> 65

<211> 21

<212> DNA

<213> sonde 2/série 1 (28588-28608)

<400> 65
gccaccgtgc tacaacttcc t 21

<210> 66

<211> 23

<212> DNA

<213> sonde 1/série 2 /sonde N/FL (28541-28563)

<400> 66
atacacccaa agaccacatt ggc 23

<210> 67

<211> 25

<212> DNA

<213> sonde 2/série 2/sonde SARS/N/LC705 (28565-28589)

<400> 67
cccgcaatcc taataacaat gctgc 25

<210> 68

<211> 30

<212> DNA

<213> artificial sequence

S226CAS108.ST25

<220>

<223> amorce ancre 14T

<400> 68

agatgaattc ggtacctttt tttttttttt

30

<210> 69

<211> 13

<212> PRT

<213> artificial sequence

<220>

<223> peptide M2-14

<400> 69

Ala Asp Asn Gly Thr Ile Thr Val Glu Glu Leu Lys Gln
1 5 10

<210> 70

<211> 12

<212> PRT

<213> artificial sequence

<220>

<223> peptide E1-12

<400> 70

Met Tyr Ser Phe Val Ser Glu Glu Thr Gly Thr Leu
1 5 10

<210> 71

<211> 24

<212> PRT

<213> artificial sequence

<220>

<223> peptide E53-72

<400> 71

Lys Pro Thr Val Tyr Val Tyr Ser Arg Val Lys Asn Leu Asn Ser Ser
1 5 10 15

S226CAS108.ST25

Glu Gly Val Pro Asp Leu Leu Val
20

<210> 72

<211> 153

<212> DNA

<213> CORONAVIRUS

<400> 72

gatattaggt ttttacctac ccaggaaaag ccaaccaacc tcgatctctt gtagatctgt 60

tctctaaacg aacttttaaaa tctgtgtagc tgtcgcctcg ctgcatgcct agtgcaccta 120

cgcagtataa acaataataa attttactgt cgt 153

<210> 73

<211> 410

<212> DNA

<213> CORONAVIRUS

<400> 73

ttctccagac aacttcaaaa ttccatgagt ggagcttctg ctgattcaac tcaggcataa 60

acactcatga tgaccacaca aggcatgagg gctatgtaaa cgttttcgca attccgttta 120

cgatacatag tctactcttg tgcagaatga attctcgtaa ctaaacagca caagtagggt 180

tagttaactt taatctcaca tagcaatctt taatcaatgt gtaacattag ggaggacttg 240

aaagagccac cacattttca tcgaggccac gcggagtagc atcgagggta cagtgaataa 300

tgctagggag agctgcctat atggaagagc cctaattgtgt aaaattaatt ttagtagtgc 360

tatcccatg tgattttaat agcttcttag gagaatgaca aaaaaaaaaa 410

<210> 74

<211> 4382

<212> PRT

<213> CORONAVIRUS

<400> 74

Met Glu Ser Leu Val Leu Gly Val Asn Glu Lys Thr His Val Gln Leu
1 5 10 15

Ser Leu Pro Val Leu Gln Val Arg Asp Val Leu Val Arg Gly Phe Gly
20 25 30

S226CAS108.ST25

Asp Ser Val₃₅ Glu Glu Ala Leu Ser₄₀ Glu Ala Arg Glu His₄₅ Leu Lys Asn
 Gly Thr₅₀ Cys Gly Leu Val₅₅ Glu Leu Glu Lys Gly Val₆₀ Leu Pro Gln Leu
 Glu Gln Pro Tyr Val₇₀ Phe Ile Lys Arg Ser₇₅ Asp Ala Leu Ser Thr₈₀ Asn
 His Gly His Lys₈₅ Val Val Glu Leu Val₉₀ Ala Glu Met Asp Gly Ile₉₅ Gln
 Tyr Gly Arg Ser₁₀₀ Gly Ile Thr Leu Gly₁₀₅ Val Leu Val Pro His₁₁₀ Val Gly
 Glu Thr₁₁₅ Pro Ile Ala Tyr Arg Asn₁₂₀ Val Leu Leu Arg Lys₁₂₅ Asn Gly Asn
 Lys Gly₁₃₀ Ala Gly Gly His Ser₁₃₅ Tyr Gly Ile Asp Leu₁₄₀ Lys Ser Tyr Asp
 Leu Gly Asp Glu Leu Gly₁₅₀ Thr Asp Pro Ile Glu₁₅₅ Asp Tyr Glu Gln Asn₁₆₀
 Trp Asn Thr Lys₁₆₅ His Gly Ser Gly Ala Leu₁₇₀ Arg Glu Leu Thr Arg₁₇₅ Glu
 Leu Asn Gly Gly₁₈₀ Ala Val Thr Arg Tyr₁₈₅ Val Asp Asn Asn Phe Cys Gly
 Pro Asp Gly₁₉₅ Tyr Pro Leu Asp Cys₂₀₀ Ile Lys Asp Phe Leu₂₀₅ Ala Arg Ala
 Gly Lys₂₁₀ Ser Met Cys Thr Leu₂₁₅ Ser Glu Gln Leu Asp₂₂₀ Tyr Ile Glu Ser
 Lys Arg Gly Val Tyr Cys₂₃₀ Cys Arg Asp His Glu₂₃₅ His Glu Ile Ala Trp₂₄₀
 Phe Thr Glu Arg Ser₂₄₅ Asp Lys Ser Tyr Glu₂₅₀ His Gln Thr Pro Phe₂₅₅ Glu
 Ile Lys Ser Ala₂₆₀ Lys Lys Phe Asp Thr₂₆₅ Phe Lys Gly Glu Cys₂₇₀ Pro Lys
 Phe Val Phe₂₇₅ Pro Leu Asn Ser Lys₂₈₀ Val Lys Val Ile Gln₂₈₅ Pro Arg Val
 Glu Lys₂₉₀ Lys Lys Thr Glu Gly₂₉₅ Phe Met Gly Arg Ile₃₀₀ Arg Ser Val Tyr

S226CAS108.ST25

Pro Val Ala Ser Pro Gln Glu Cys Asn Asn Met His Leu Ser Thr Leu
 305 310 315 320
 Met Lys Cys Asn His Cys Asp Glu Val Ser Trp Gln Thr Cys Asp Phe
 325 330 335
 Leu Lys Ala Thr Cys Glu His Cys Gly Thr Glu Asn Leu Val Ile Glu
 340 345 350
 Gly Pro Thr Thr Cys Gly Tyr Leu Pro Thr Asn Ala Val Val Lys Met
 355 360 365
 Pro Cys Pro Ala Cys Gln Asp Pro Glu Ile Gly Pro Glu His Ser Val
 370 375 380
 Ala Asp Tyr His Asn His Ser Asn Ile Glu Thr Arg Leu Arg Lys Gly
 385 390 395 400
 Gly Arg Thr Arg Cys Phe Gly Gly Cys Val Phe Ala Tyr Val Gly Cys
 405 410 415
 Tyr Asn Lys Arg Ala Tyr Trp Val Pro Arg Ala Ser Ala Asp Ile Gly
 420 425 430
 Ser Gly His Thr Gly Ile Thr Gly Asp Asn Val Glu Thr Leu Asn Glu
 435 440 445
 Asp Leu Leu Glu Ile Leu Ser Arg Glu Arg Val Asn Ile Asn Ile Val
 450 455 460
 Gly Asp Phe His Leu Asn Glu Glu Val Ala Ile Ile Leu Ala Ser Phe
 465 470 475 480
 Ser Ala Ser Thr Ser Ala Phe Ile Asp Thr Ile Lys Ser Leu Asp Tyr
 485 490 495
 Lys Ser Phe Lys Thr Ile Val Glu Ser Cys Gly Asn Tyr Lys Val Thr
 500 505 510
 Lys Gly Lys Pro Val Lys Gly Ala Trp Asn Ile Gly Gln Gln Arg Ser
 515 520 525
 Val Leu Thr Pro Leu Cys Gly Phe Pro Ser Gln Ala Ala Gly Val Ile
 530 535 540
 Arg Ser Ile Phe Ala Arg Thr Leu Asp Ala Ala Asn His Ser Ile Pro
 545 550 555 560
 Asp Leu Gln Arg Ala Ala Val Thr Ile Leu Asp Gly Ile Ser Glu Gln
 565 570 575

S226CAS108.ST25

Ser Leu Arg Leu Val Asp Ala Met Val Tyr Thr Ser Asp Leu Leu Thr
 580 585 590
 Asn Ser Val Ile Ile Met Ala Tyr Val Thr Gly Gly Leu Val Gln Gln
 595 600 605
 Thr Ser Gln Trp Leu Ser Asn Leu Leu Gly Thr Thr Val Glu Lys Leu
 610 615 620
 Arg Pro Ile Phe Glu Trp Ile Glu Ala Lys Leu Ser Ala Gly Val Glu
 625 630 635 640
 Phe Leu Lys Asp Ala Trp Glu Ile Leu Lys Phe Leu Ile Thr Gly Val
 645 650 655
 Phe Asp Ile Val Lys Gly Gln Ile Gln Val Ala Ser Asp Asn Ile Lys
 660 665 670
 Asp Cys Val Lys Cys Phe Ile Asp Val Val Asn Lys Ala Leu Glu Met
 675 680 685
 Cys Ile Asp Gln Val Thr Ile Ala Gly Ala Lys Leu Arg Ser Leu Asn
 690 695 700
 Leu Gly Glu Val Phe Ile Ala Gln Ser Lys Gly Leu Tyr Arg Gln Cys
 705 710 715 720
 Ile Arg Gly Lys Glu Gln Leu Gln Leu Leu Met Pro Leu Lys Ala Pro
 725 730 735
 Lys Glu Val Thr Phe Leu Glu Gly Asp Ser His Asp Thr Val Leu Thr
 740 745 750
 Ser Glu Glu Val Val Leu Lys Asn Gly Glu Leu Glu Ala Leu Glu Thr
 755 760 765
 Pro Val Asp Ser Phe Thr Asn Gly Ala Ile Val Gly Thr Pro Val Cys
 770 775 780
 Val Asn Gly Leu Met Leu Leu Glu Ile Lys Asp Lys Glu Gln Tyr Cys
 785 790 795 800
 Ala Leu Ser Pro Gly Leu Leu Ala Thr Asn Asn Val Phe Arg Leu Lys
 805 810 815
 Gly Gly Ala Pro Ile Lys Gly Val Thr Phe Gly Glu Asp Thr Val Trp
 820 825 830
 Glu Val Gln Gly Tyr Lys Asn Val Arg Ile Thr Phe Glu Leu Asp Glu
 835 840 845

S226CAS108.ST25

Arg Val Asp Lys Val Leu Asn Glu Lys Cys Ser Val Tyr Thr Val Glu
 850 855 860
 Ser Gly Thr Glu Val Thr Glu Phe Ala Cys Val Val Ala Glu Ala Val
 865 870 875 880
 Val Lys Thr Leu Gln Pro Val Ser Asp Leu Leu Thr Asn Met Gly Ile
 885 890 895
 Asp Leu Asp Glu Trp Ser Val Ala Thr Phe Tyr Leu Phe Asp Asp Ala
 900 905 910
 Gly Glu Glu Asn Phe Ser Ser Arg Met Tyr Cys Ser Phe Tyr Pro Pro
 915 920 925
 Asp Glu Glu Glu Glu Asp Asp Ala Glu Cys Glu Glu Glu Glu Ile Asp
 930 935 940
 Glu Thr Cys Glu His Glu Tyr Gly Thr Glu Asp Asp Tyr Gln Gly Leu
 945 950 955 960
 Pro Leu Glu Phe Gly Ala Ser Ala Glu Thr Val Arg Val Glu Glu Glu
 965 970 975
 Glu Glu Glu Asp Trp Leu Asp Asp Thr Thr Glu Gln Ser Glu Ile Glu
 980 985 990
 Pro Glu Pro Glu Pro Thr Pro Glu Glu Pro Val Asn Gln Phe Thr Gly
 995 1000 1005
 Tyr Leu Lys Leu Thr Asp Asn Val Ala Ile Lys Cys Val Asp Ile
 1010 1015 1020
 Val Lys Glu Ala Gln Ser Ala Asn Pro Met Val Ile Val Asn Ala
 1025 1030 1035
 Ala Asn Ile His Leu Lys His Gly Gly Gly Val Ala Gly Ala Leu
 1040 1045 1050
 Asn Lys Ala Thr Asn Gly Ala Met Gln Lys Glu Ser Asp Asp Tyr
 1055 1060 1065
 Ile Lys Leu Asn Gly Pro Leu Thr Val Gly Gly Ser Cys Leu Leu
 1070 1075 1080
 Ser Gly His Asn Leu Ala Lys Lys Cys Leu His Val Val Gly Pro
 1085 1090 1095
 Asn Leu Asn Ala Gly Glu Asp Ile Gln Leu Leu Lys Ala Ala Tyr
 1100 1105 1110

S226CAS108.ST25

Glu Asn Phe Asn Ser Gln Asp Ile Leu Leu Ala Pro Leu Leu Ser
 1115 1120 1125
 Ala Gly Ile Phe Gly Ala Lys Pro Leu Gln Ser Leu Gln Val Cys
 1130 1135 1140
 Val Gln Thr Val Arg Thr Gln Val Tyr Ile Ala Val Asn Asp Lys
 1145 1150 1155
 Ala Leu Tyr Glu Gln Val Val Met Asp Tyr Leu Asp Asn Leu Lys
 1160 1165 1170
 Pro Arg Val Glu Ala Pro Lys Gln Glu Glu Pro Pro Asn Thr Glu
 1175 1180 1185
 Asp Ser Lys Thr Glu Glu Lys Ser Val Val Gln Lys Pro Val Asp
 1190 1195 1200
 Val Lys Pro Lys Ile Lys Ala Cys Ile Asp Glu Val Thr Thr Thr
 1205 1210 1215
 Leu Glu Glu Thr Lys Phe Leu Thr Asn Lys Leu Leu Leu Phe Ala
 1220 1225 1230
 Asp Ile Asn Gly Lys Leu Tyr His Asp Ser Gln Asn Met Leu Arg
 1235 1240 1245
 Gly Glu Asp Met Ser Phe Leu Glu Lys Asp Ala Pro Tyr Met Val
 1250 1255 1260
 Gly Asp Val Ile Thr Ser Gly Asp Ile Thr Cys Val Val Ile Pro
 1265 1270 1275
 Ser Lys Lys Ala Gly Gly Thr Thr Glu Met Leu Ser Arg Ala Leu
 1280 1285 1290
 Lys Lys Val Pro Val Asp Glu Tyr Ile Thr Thr Tyr Pro Gly Gln
 1295 1300 1305
 Gly Cys Ala Gly Tyr Thr Leu Glu Glu Ala Lys Thr Ala Leu Lys
 1310 1315 1320
 Lys Cys Lys Ser Ala Phe Tyr Val Leu Pro Ser Glu Ala Pro Asn
 1325 1330 1335
 Ala Lys Glu Glu Ile Leu Gly Thr Val Ser Trp Asn Leu Arg Glu
 1340 1345 1350
 Met Leu Ala His Ala Glu Glu Thr Arg Lys Leu Met Pro Ile Cys
 1355 1360 1365

S226CAS108.ST25

Met Asp Val Arg Ala Ile Met Ala Thr Ile Gln Arg Lys Tyr Lys
 1370 1375 1380
 Gly Ile Lys Ile Gln Glu Gly Ile Val Asp Tyr Gly Val Arg Phe
 1385 1390 1395
 Phe Phe Tyr Thr Ser Lys Glu Pro Val Ala Ser Ile Ile Thr Lys
 1400 1405 1410
 Leu Asn Ser Leu Asn Glu Pro Leu Val Thr Met Pro Ile Gly Tyr
 1415 1420 1425
 Val Thr His Gly Phe Asn Leu Glu Glu Ala Ala Arg Cys Met Arg
 1430 1435 1440
 Ser Leu Lys Ala Pro Ala Val Val Ser Val Ser Ser Pro Asp Ala
 1445 1450 1455
 Val Thr Thr Tyr Asn Gly Tyr Leu Thr Ser Ser Ser Lys Thr Ser
 1460 1465 1470
 Glu Glu His Phe Val Glu Thr Val Ser Leu Ala Gly Ser Tyr Arg
 1475 1480 1485
 Asp Trp Ser Tyr Ser Gly Gln Arg Thr Glu Leu Gly Val Glu Phe
 1490 1495 1500
 Leu Lys Arg Gly Asp Lys Ile Val Tyr His Thr Leu Glu Ser Pro
 1505 1510 1515
 Val Glu Phe His Leu Asp Gly Glu Val Leu Ser Leu Asp Lys Leu
 1520 1525 1530
 Lys Ser Leu Leu Ser Leu Arg Glu Val Lys Thr Ile Lys Val Phe
 1535 1540 1545
 Thr Thr Val Asp Asn Thr Asn Leu His Thr Gln Leu Val Asp Met
 1550 1555 1560
 Ser Met Thr Tyr Gly Gln Gln Phe Gly Pro Thr Tyr Leu Asp Gly
 1565 1570 1575
 Ala Asp Val Thr Lys Ile Lys Pro His Val Asn His Glu Gly Lys
 1580 1585 1590
 Thr Phe Phe Val Leu Pro Ser Asp Asp Thr Leu Arg Ser Glu Ala
 1595 1600 1605
 Phe Glu Tyr Tyr His Thr Leu Asp Glu Ser Phe Leu Gly Arg Tyr
 1610 1615 1620

S226CAS108.ST25

Met	Ser	Ala	Leu	Asn	His	Thr	Lys	Lys	Trp	Lys	Phe	Pro	Gln	Val
	1625					1630					1635			
Gly	Gly	Leu	Thr	Ser	Ile	Lys	Trp	Ala	Asp	Asn	Asn	Cys	Tyr	Leu
	1640					1645					1650			
Ser	Ser	Val	Leu	Leu	Ala	Leu	Gln	Gln	Leu	Glu	Val	Lys	Phe	Asn
	1655					1660					1665			
Ala	Pro	Ala	Leu	Gln	Glu	Ala	Tyr	Tyr	Arg	Ala	Arg	Ala	Gly	Asp
	1670					1675					1680			
Ala	Ala	Asn	Phe	Cys	Ala	Leu	Ile	Leu	Ala	Tyr	Ser	Asn	Lys	Thr
	1685					1690					1695			
Val	Gly	Glu	Leu	Gly	Asp	Val	Arg	Glu	Thr	Met	Thr	His	Leu	Leu
	1700					1705					1710			
Gln	His	Ala	Asn	Leu	Glu	Ser	Ala	Lys	Arg	Val	Leu	Asn	Val	Val
	1715					1720					1725			
Cys	Lys	His	Cys	Gly	Gln	Lys	Thr	Thr	Thr	Leu	Thr	Gly	Val	Glu
	1730					1735					1740			
Ala	Val	Met	Tyr	Met	Gly	Thr	Leu	Ser	Tyr	Asp	Asn	Leu	Lys	Thr
	1745					1750					1755			
Gly	Val	Ser	Ile	Pro	Cys	Val	Cys	Gly	Arg	Asp	Ala	Thr	Gln	Tyr
	1760					1765					1770			
Leu	Val	Gln	Gln	Glu	Ser	Ser	Phe	Val	Met	Met	Ser	Ala	Pro	Pro
	1775					1780					1785			
Ala	Glu	Tyr	Lys	Leu	Gln	Gln	Gly	Thr	Phe	Leu	Cys	Ala	Asn	Glu
	1790					1795					1800			
Tyr	Thr	Gly	Asn	Tyr	Gln	Cys	Gly	His	Tyr	Thr	His	Ile	Thr	Ala
	1805					1810					1815			
Lys	Glu	Thr	Leu	Tyr	Arg	Ile	Asp	Gly	Ala	His	Leu	Thr	Lys	Met
	1820					1825					1830			
Ser	Glu	Tyr	Lys	Gly	Pro	Val	Thr	Asp	Val	Phe	Tyr	Lys	Glu	Thr
	1835					1840					1845			
Ser	Tyr	Thr	Thr	Thr	Ile	Lys	Pro	Val	Ser	Tyr	Lys	Leu	Asp	Gly
	1850					1855					1860			
Val	Thr	Tyr	Thr	Glu	Ile	Glu	Pro	Lys	Leu	Asp	Gly	Tyr	Tyr	Lys
	1865					1870					1875			

S226CAS108.ST25

Lys Asp Asn Ala Tyr Tyr Thr Glu Gln Pro Ile Asp Leu Val Pro
 1880 1885 1890
 Thr Gln Pro Leu Pro Asn Ala Ser Phe Asp Asn Phe Lys Leu Thr
 1895 1900 1905
 Cys Ser Asn Thr Lys Phe Ala Asp Asp Leu Asn Gln Met Thr Gly
 1910 1915 1920
 Phe Thr Lys Pro Ala Ser Arg Glu Leu Ser Val Thr Phe Phe Pro
 1925 1930 1935
 Asp Leu Asn Gly Asp Val Val Ala Ile Asp Tyr Arg His Tyr Ser
 1940 1945 1950
 Ala Ser Phe Lys Lys Gly Ala Lys Leu Leu His Lys Pro Ile Val
 1955 1960 1965
 Trp His Ile Asn Gln Ala Thr Thr Lys Thr Thr Phe Lys Pro Asn
 1970 1975 1980
 Thr Trp Cys Leu Arg Cys Leu Trp Ser Thr Lys Pro Val Asp Thr
 1985 1990 1995
 Ser Asn Ser Phe Glu Val Leu Ala Val Glu Asp Thr Gln Gly Met
 2000 2005 2010
 Asp Asn Leu Ala Cys Glu Ser Gln Gln Pro Thr Ser Glu Glu Val
 2015 2020 2025
 Val Glu Asn Pro Thr Ile Gln Lys Glu Val Ile Glu Cys Asp Val
 2030 2035 2040
 Lys Thr Thr Glu Val Val Gly Asn Val Ile Leu Lys Pro Ser Asp
 2045 2050 2055
 Glu Gly Val Lys Val Thr Gln Glu Leu Gly His Glu Asp Leu Met
 2060 2065 2070
 Ala Ala Tyr Val Glu Asn Thr Ser Ile Thr Ile Lys Lys Pro Asn
 2075 2080 2085
 Glu Leu Ser Leu Ala Leu Gly Leu Lys Thr Ile Ala Thr His Gly
 2090 2095 2100
 Ile Ala Ala Ile Asn Ser Val Pro Trp Ser Lys Ile Leu Ala Tyr
 2105 2110 2115
 Val Lys Pro Phe Leu Gly Gln Ala Ala Ile Thr Thr Ser Asn Cys
 2120 2125 2130

S226CAS108.ST25

Ala Lys Arg Leu Ala Gln Arg Val Phe Asn Asn Tyr Met Pro Tyr
 2135 2140 2145
 Val Phe Thr Leu Leu Phe Gln Leu Cys Thr Phe Thr Lys Ser Thr
 2150 2155 2160
 Asn Ser Arg Ile Arg Ala Ser Leu Pro Thr Thr Ile Ala Lys Asn
 2165 2170 2175
 Ser Val Lys Ser Val Ala Lys Leu Cys Leu Asp Ala Gly Ile Asn
 2180 2185 2190
 Tyr Val Lys Ser Pro Lys Phe Ser Lys Leu Phe Thr Ile Ala Met
 2195 2200 2205
 Trp Leu Leu Leu Leu Ser Ile Cys Leu Gly Ser Leu Ile Cys Val
 2210 2215 2220
 Thr Ala Ala Phe Gly Val Leu Leu Ser Asn Phe Gly Ala Pro Ser
 2225 2230 2235
 Tyr Cys Asn Gly Val Arg Glu Leu Tyr Leu Asn Ser Ser Asn Val
 2240 2245 2250
 Thr Thr Met Asp Phe Cys Glu Gly Ser Phe Pro Cys Ser Ile Cys
 2255 2260 2265
 Leu Ser Gly Leu Asp Ser Leu Asp Ser Tyr Pro Ala Leu Glu Thr
 2270 2275 2280
 Ile Gln Val Thr Ile Ser Ser Tyr Lys Leu Asp Leu Thr Ile Leu
 2285 2290 2295
 Gly Leu Ala Ala Glu Trp Val Leu Ala Tyr Met Leu Phe Thr Lys
 2300 2305 2310
 Phe Phe Tyr Leu Leu Gly Leu Ser Ala Ile Met Gln Val Phe Phe
 2315 2320 2325
 Gly Tyr Phe Ala Ser His Phe Ile Ser Asn Ser Trp Leu Met Trp
 2330 2335 2340
 Phe Ile Ile Ser Ile Val Gln Met Ala Pro Val Ser Ala Met Val
 2345 2350 2355
 Arg Met Tyr Ile Phe Phe Ala Ser Phe Tyr Tyr Ile Trp Lys Ser
 2360 2365 2370
 Tyr Val His Ile Met Asp Gly Cys Thr Ser Ser Thr Cys Met Met
 2375 2380 2385

S226CAS108.ST25

Cys Tyr Lys Arg Asn Arg Ala Thr Arg Val Glu Cys Thr Thr Ile
 2390 2395 2400
 Val Asn Gly Met Lys Arg Ser Phe Tyr Val Tyr Ala Asn Gly Gly
 2405 2410 2415
 Arg Gly Phe Cys Lys Thr His Asn Trp Asn Cys Leu Asn Cys Asp
 2420 2425 2430
 Thr Phe Cys Thr Gly Ser Thr Phe Ile Ser Asp Glu Val Ala Arg
 2435 2440 2445
 Asp Leu Ser Leu Gln Phe Lys Arg Pro Ile Asn Pro Thr Asp Gln
 2450 2455 2460
 Ser Ser Tyr Ile Val Asp Ser Val Ala Val Lys Asn Gly Ala Leu
 2465 2470 2475
 His Leu Tyr Phe Asp Lys Ala Gly Gln Lys Thr Tyr Glu Arg His
 2480 2485 2490
 Pro Leu Ser His Phe Val Asn Leu Asp Asn Leu Arg Ala Asn Asn
 2495 2500 2505
 Thr Lys Gly Ser Leu Pro Ile Asn Val Ile Val Phe Asp Gly Lys
 2510 2515 2520
 Ser Lys Cys Asp Glu Ser Ala Ser Lys Ser Ala Ser Val Tyr Tyr
 2525 2530 2535
 Ser Gln Leu Met Cys Gln Pro Ile Leu Leu Leu Asp Gln Ala Leu
 2540 2545 2550
 Val Ser Asp Val Gly Asp Ser Thr Glu Val Ser Val Lys Met Phe
 2555 2560 2565
 Asp Ala Tyr Val Asp Thr Phe Ser Ala Thr Phe Ser Val Pro Met
 2570 2575 2580
 Glu Lys Leu Lys Ala Leu Val Ala Thr Ala His Ser Glu Leu Ala
 2585 2590 2595
 Lys Gly Val Ala Leu Asp Gly Val Leu Ser Thr Phe Val Ser Ala
 2600 2605 2610
 Ala Arg Gln Gly Val Val Asp Thr Asp Val Asp Thr Lys Asp Val
 2615 2620 2625
 Ile Glu Cys Leu Lys Leu Ser His His Ser Asp Leu Glu Val Thr
 2630 2635 2640

S226CAS108.ST25

Gly Asp Ser Cys Asn Asn Phe Met Leu Thr Tyr Asn Lys Val Glu
 2645 2650 2655
 Asn Met Thr Pro Arg Asp Leu Gly Ala Cys Ile Asp Cys Asn Ala
 2660 2665 2670
 Arg His Ile Asn Ala Gln Val Ala Lys Ser His Asn Val Ser Leu
 2675 2680 2685
 Ile Trp Asn Val Lys Asp Tyr Met Ser Leu Ser Glu Gln Leu Arg
 2690 2695 2700
 Lys Gln Ile Arg Ser Ala Ala Lys Lys Asn Asn Ile Pro Phe Arg
 2705 2710 2715
 Leu Thr Cys Ala Thr Thr Arg Gln Val Val Asn Val Ile Thr Thr
 2720 2725 2730
 Lys Ile Ser Leu Lys Gly Gly Lys Ile Val Ser Thr Cys Phe Lys
 2735 2740 2745
 Leu Met Leu Lys Ala Thr Leu Leu Cys Val Leu Ala Ala Leu Val
 2750 2755 2760
 Cys Tyr Ile Val Met Pro Val His Thr Leu Ser Ile His Asp Gly
 2765 2770 2775
 Tyr Thr Asn Glu Ile Ile Gly Tyr Lys Ala Ile Gln Asp Gly Val
 2780 2785 2790
 Thr Arg Asp Ile Ile Ser Thr Asp Asp Cys Phe Ala Asn Lys His
 2795 2800 2805
 Ala Gly Phe Asp Ala Trp Phe Ser Gln Arg Gly Gly Ser Tyr Lys
 2810 2815 2820
 Asn Asp Lys Ser Cys Pro Val Val Ala Ala Ile Ile Thr Arg Glu
 2825 2830 2835
 Ile Gly Phe Ile Val Pro Gly Leu Pro Gly Thr Val Leu Arg Ala
 2840 2845 2850
 Ile Asn Gly Asp Phe Leu His Phe Leu Pro Arg Val Phe Ser Ala
 2855 2860 2865
 Val Gly Asn Ile Cys Tyr Thr Pro Ser Lys Leu Ile Glu Tyr Ser
 2870 2875 2880
 Asp Phe Ala Thr Ser Ala Cys Val Leu Ala Ala Glu Cys Thr Ile
 2885 2890 2895

S226CAS108.ST25

Phe Lys Asp Ala Met Gly Lys Pro Val Pro Tyr Cys Tyr Asp Thr
 2900 2905 2910
 Asn Leu Leu Glu Gly Ser Ile Ser Tyr Ser Glu Leu Arg Pro Asp
 2915 2920 2925
 Thr Arg Tyr Val Leu Met Asp Gly Ser Ile Ile Gln Phe Pro Asn
 2930 2935 2940
 Thr Tyr Leu Glu Gly Ser Val Arg Val Val Thr Thr Phe Asp Ala
 2945 2950 2955
 Glu Tyr Cys Arg His Gly Thr Cys Glu Arg Ser Glu Val Gly Ile
 2960 2965 2970
 Cys Leu Ser Thr Ser Gly Arg Trp Val Leu Asn Asn Glu His Tyr
 2975 2980 2985
 Arg Ala Leu Ser Gly Val Phe Cys Gly Val Asp Ala Met Asn Leu
 2990 2995 3000
 Ile Ala Asn Ile Phe Thr Pro Leu Val Gln Pro Val Gly Ala Leu
 3005 3010 3015
 Asp Val Ser Ala Ser Val Val Ala Gly Gly Ile Ile Ala Ile Leu
 3020 3025 3030
 Val Thr Cys Ala Ala Tyr Tyr Phe Met Lys Phe Arg Arg Val Phe
 3035 3040 3045
 Gly Glu Tyr Asn His Val Val Ala Ala Asn Ala Leu Leu Phe Leu
 3050 3055 3060
 Met Ser Phe Thr Ile Leu Cys Leu Val Pro Ala Tyr Ser Phe Leu
 3065 3070 3075
 Pro Gly Val Tyr Ser Val Phe Tyr Leu Tyr Leu Thr Phe Tyr Phe
 3080 3085 3090
 Thr Asn Asp Val Ser Phe Leu Ala His Leu Gln Trp Phe Ala Met
 3095 3100 3105
 Phe Ser Pro Ile Val Pro Phe Trp Ile Thr Ala Ile Tyr Val Phe
 3110 3115 3120
 Cys Ile Ser Leu Lys His Cys His Trp Phe Phe Asn Asn Tyr Leu
 3125 3130 3135
 Arg Lys Arg Val Met Phe Asn Gly Val Thr Phe Ser Thr Phe Glu
 3140 3145 3150

S226CAS108.ST25

Glu Ala Ala Leu Cys Thr Phe Leu Leu Asn Lys Glu Met Tyr Leu
 3155 3160 3165
 Lys Leu Arg Ser Glu Thr Leu Leu Pro Leu Thr Gln Tyr Asn Arg
 3170 3175 3180
 Tyr Leu Ala Leu Tyr Asn Lys Tyr Lys Tyr Phe Ser Gly Ala Leu
 3185 3190 3195
 Asp Thr Thr Ser Tyr Arg Glu Ala Ala Cys Cys His Leu Ala Lys
 3200 3205 3210
 Ala Leu Asn Asp Phe Ser Asn Ser Gly Ala Asp Val Leu Tyr Gln
 3215 3220 3225
 Pro Pro Gln Thr Ser Ile Thr Ser Ala Val Leu Gln Ser Gly Phe
 3230 3235 3240
 Arg Lys Met Ala Phe Pro Ser Gly Lys Val Glu Gly Cys Met Val
 3245 3250 3255
 Gln Val Thr Cys Gly Thr Thr Thr Leu Asn Gly Leu Trp Leu Asp
 3260 3265 3270
 Asp Thr Val Tyr Cys Pro Arg His Val Ile Cys Thr Ala Glu Asp
 3275 3280 3285
 Met Leu Asn Pro Asn Tyr Glu Asp Leu Leu Ile Arg Lys Ser Asn
 3290 3295 3300
 His Ser Phe Leu Val Gln Ala Gly Asn Val Gln Leu Arg Val Ile
 3305 3310 3315
 Gly His Ser Met Gln Asn Cys Leu Leu Arg Leu Lys Val Asp Thr
 3320 3325 3330
 Ser Asn Pro Lys Thr Pro Lys Tyr Lys Phe Val Arg Ile Gln Pro
 3335 3340 3345
 Gly Gln Thr Phe Ser Val Leu Ala Cys Tyr Asn Gly Ser Pro Ser
 3350 3355 3360
 Gly Val Tyr Gln Cys Ala Met Arg Pro Asn His Thr Ile Lys Gly
 3365 3370 3375
 Ser Phe Leu Asn Gly Ser Cys Gly Ser Val Gly Phe Asn Ile Asp
 3380 3385 3390
 Tyr Asp Cys Val Ser Phe Cys Tyr Met His His Met Glu Leu Pro
 3395 3400 3405

S226CAS108.ST25

Thr Gly Val His Ala Gly Thr Asp Leu Glu Gly Lys Phe Tyr Gly
 3410 3415 3420
 Pro Phe Val Asp Arg Gln Thr Ala Gln Ala Ala Gly Thr Asp Thr
 3425 3430 3435
 Thr Ile Thr Leu Asn Val Leu Ala Trp Leu Tyr Ala Ala Val Ile
 3440 3445 3450
 Asn Gly Asp Arg Trp Phe Leu Asn Arg Phe Thr Thr Thr Leu Asn
 3455 3460 3465
 Asp Phe Asn Leu Val Ala Met Lys Tyr Asn Tyr Glu Pro Leu Thr
 3470 3475 3480
 Gln Asp His Val Asp Ile Leu Gly Pro Leu Ser Ala Gln Thr Gly
 3485 3490 3495
 Ile Ala Val Leu Asp Met Cys Ala Ala Leu Lys Glu Leu Leu Gln
 3500 3505 3510
 Asn Gly Met Asn Gly Arg Thr Ile Leu Gly Ser Thr Ile Leu Glu
 3515 3520 3525
 Asp Glu Phe Thr Pro Phe Asp Val Val Arg Gln Cys Ser Gly Val
 3530 3535 3540
 Thr Phe Gln Gly Lys Phe Lys Lys Ile Val Lys Gly Thr His His
 3545 3550 3555
 Trp Met Leu Leu Thr Phe Leu Thr Ser Leu Leu Ile Leu Val Gln
 3560 3565 3570
 Ser Thr Gln Trp Ser Leu Phe Phe Phe Val Tyr Glu Asn Ala Phe
 3575 3580 3585
 Leu Pro Phe Thr Leu Gly Ile Met Ala Ile Ala Ala Cys Ala Met
 3590 3595 3600
 Leu Leu Val Lys His Lys His Ala Phe Leu Cys Leu Phe Leu Leu
 3605 3610 3615
 Pro Ser Leu Ala Thr Val Ala Tyr Phe Asn Met Val Tyr Met Pro
 3620 3625 3630
 Ala Ser Trp Val Met Arg Ile Met Thr Trp Leu Glu Leu Ala Asp
 3635 3640 3645
 Thr Ser Leu Ser Gly Tyr Arg Leu Lys Asp Cys Val Met Tyr Ala
 3650 3655 3660

S226CAS108.ST25

Ser Ala Leu Val Leu Leu Ile Leu Met Thr Ala Arg Thr Val Tyr
 3665 3670 3675
 Asp Asp Ala Ala Arg Arg Val Trp Thr Leu Met Asn Val Ile Thr
 3680 3685 3690
 Leu Val Tyr Lys Val Tyr Tyr Gly Asn Ala Leu Asp Gln Ala Ile
 3695 3700 3705
 Ser Met Trp Ala Leu Val Ile Ser Val Thr Ser Asn Tyr Ser Gly
 3710 3715 3720
 Val Val Thr Thr Ile Met Phe Leu Ala Arg Ala Ile Val Phe Val
 3725 3730 3735
 Cys Val Glu Tyr Tyr Pro Leu Leu Phe Ile Thr Gly Asn Thr Leu
 3740 3745 3750
 Gln Cys Ile Met Leu Val Tyr Cys Phe Leu Gly Tyr Cys Cys Cys
 3755 3760 3765
 Cys Tyr Phe Gly Leu Phe Cys Leu Leu Asn Arg Tyr Phe Arg Leu
 3770 3775 3780
 Thr Leu Gly Val Tyr Asp Tyr Leu Val Ser Thr Gln Glu Phe Arg
 3785 3790 3795
 Tyr Met Asn Ser Gln Gly Leu Leu Pro Pro Lys Ser Ser Ile Asp
 3800 3805 3810
 Ala Phe Lys Leu Asn Ile Lys Leu Leu Gly Ile Gly Gly Lys Pro
 3815 3820 3825
 Cys Ile Lys Val Ala Thr Val Gln Ser Lys Met Ser Asp Val Lys
 3830 3835 3840
 Cys Thr Ser Val Val Leu Leu Ser Val Leu Gln Gln Leu Arg Val
 3845 3850 3855
 Glu Ser Ser Ser Lys Leu Trp Ala Gln Cys Val Gln Leu His Asn
 3860 3865 3870
 Asp Ile Leu Leu Ala Lys Asp Thr Thr Glu Ala Phe Glu Lys Met
 3875 3880 3885
 Val Ser Leu Leu Ser Val Leu Leu Ser Met Gln Gly Ala Val Asp
 3890 3895 3900
 Ile Asn Arg Leu Cys Glu Glu Met Leu Asp Asn Arg Ala Thr Leu
 3905 3910 3915

S226CAS108.ST25

Gln Ala Ile Ala Ser Glu Phe Ser Ser Leu Pro Ser Tyr Ala Ala
 3920 3925 3930
 Tyr Ala Thr Ala Gln Glu Ala Tyr Glu Gln Ala Val Ala Asn Gly
 3935 3940 3945
 Asp Ser Glu Val Val Leu Lys Lys Leu Lys Lys Ser Leu Asn Val
 3950 3955 3960
 Ala Lys Ser Glu Phe Asp Arg Asp Ala Ala Met Gln Arg Lys Leu
 3965 3970 3975
 Glu Lys Met Ala Asp Gln Ala Met Thr Gln Met Tyr Lys Gln Ala
 3980 3985 3990
 Arg Ser Glu Asp Lys Arg Ala Lys Val Thr Ser Ala Met Gln Thr
 3995 4000 4005
 Met Leu Phe Thr Met Leu Arg Lys Leu Asp Asn Asp Ala Leu Asn
 4010 4015 4020
 Asn Ile Ile Asn Asn Ala Arg Asp Gly Cys Val Pro Leu Asn Ile
 4025 4030 4035
 Ile Pro Leu Thr Thr Ala Ala Lys Leu Met Val Val Val Pro Asp
 4040 4045 4050
 Tyr Gly Thr Tyr Lys Asn Thr Cys Asp Gly Asn Thr Phe Thr Tyr
 4055 4060 4065
 Ala Ser Ala Leu Trp Glu Ile Gln Gln Val Val Asp Ala Asp Ser
 4070 4075 4080
 Lys Ile Val Gln Leu Ser Glu Ile Asn Met Asp Asn Ser Pro Asn
 4085 4090 4095
 Leu Ala Trp Pro Leu Ile Val Thr Ala Leu Arg Ala Asn Ser Ala
 4100 4105 4110
 Val Lys Leu Gln Asn Asn Glu Leu Ser Pro Val Ala Leu Arg Gln
 4115 4120 4125
 Met Ser Cys Ala Ala Gly Thr Thr Gln Thr Ala Cys Thr Asp Asp
 4130 4135 4140
 Asn Ala Leu Ala Tyr Tyr Asn Asn Ser Lys Gly Gly Arg Phe Val
 4145 4150 4155
 Leu Ala Leu Leu Ser Asp His Gln Asp Leu Lys Trp Ala Arg Phe
 4160 4165 4170

S226CAS108.ST25

Pro Lys Ser Asp Gly Thr Gly Thr Ile Tyr Thr Glu Leu Glu Pro
 4175 4180 4185
 Pro Cys Arg Phe Val Thr Asp Thr Pro Lys Gly Pro Lys Val Lys
 4190 4195 4200
 Tyr Leu Tyr Phe Ile Lys Gly Leu Asn Asn Leu Asn Arg Gly Met
 4205 4210 4215
 Val Leu Gly Ser Leu Ala Ala Thr Val Arg Leu Gln Ala Gly Asn
 4220 4225 4230
 Ala Thr Glu Val Pro Ala Asn Ser Thr Val Leu Ser Phe Cys Ala
 4235 4240 4245
 Phe Ala Val Asp Pro Ala Lys Ala Tyr Lys Asp Tyr Leu Ala Ser
 4250 4255 4260
 Gly Gly Gln Pro Ile Thr Asn Cys Val Lys Met Leu Cys Thr His
 4265 4270 4275
 Thr Gly Thr Gly Gln Ala Ile Thr Val Thr Pro Glu Ala Asn Met
 4280 4285 4290
 Asp Gln Glu Ser Phe Gly Gly Ala Ser Cys Cys Leu Tyr Cys Arg
 4295 4300 4305
 Cys His Ile Asp His Pro Asn Pro Lys Gly Phe Cys Asp Leu Lys
 4310 4315 4320
 Gly Lys Tyr Val Gln Ile Pro Thr Thr Cys Ala Asn Asp Pro Val
 4325 4330 4335
 Gly Phe Thr Leu Arg Asn Thr Val Cys Thr Val Cys Gly Met Trp
 4340 4345 4350
 Lys Gly Tyr Gly Cys Ser Cys Asp Gln Leu Arg Glu Pro Leu Met
 4355 4360 4365
 Gln Ser Ala Asp Ala Ser Thr Phe Leu Asn Gly Phe Ala Val
 4370 4375 4380

<210> 75

<211> 2695

<212> PRT

<213> CORONAVIRUS

S226CAS108.ST25

<400> 75

Arg Val Cys Gly Val Ser Ala Ala Arg Leu Thr Pro Cys Gly Thr Gly
 1 5 10 15
 Thr Ser Thr Asp Val Val Tyr Arg Ala Phe Asp Ile Tyr Asn Glu Lys
 20 25 30
 Val Ala Gly Phe Ala Lys Phe Leu Lys Thr Asn Cys Cys Arg Phe Gln
 35 40 45
 Glu Lys Asp Glu Glu Gly Asn Leu Leu Asp Ser Tyr Phe Val Val Lys
 50 55 60
 Arg His Thr Met Ser Asn Tyr Gln His Glu Glu Thr Ile Tyr Asn Leu
 65 70 75 80
 Val Lys Asp Cys Pro Ala Val Ala Val His Asp Phe Phe Lys Phe Arg
 85 90 95
 Val Asp Gly Asp Met Val Pro His Ile Ser Arg Gln Arg Leu Thr Lys
 100 105 110
 Tyr Thr Met Ala Asp Leu Val Tyr Ala Leu Arg His Phe Asp Glu Gly
 115 120 125
 Asn Cys Asp Thr Leu Lys Glu Ile Leu Val Thr Tyr Asn Cys Cys Asp
 130 135 140
 Asp Asp Tyr Phe Asn Lys Lys Asp Trp Tyr Asp Phe Val Glu Asn Pro
 145 150 155 160
 Asp Ile Leu Arg Val Tyr Ala Asn Leu Gly Glu Arg Val Arg Gln Ser
 165 170 175
 Leu Leu Lys Thr Val Gln Phe Cys Asp Ala Met Arg Asp Ala Gly Ile
 180 185 190
 Val Gly Val Leu Thr Leu Asp Asn Gln Asp Leu Asn Gly Asn Trp Tyr
 195 200 205
 Asp Phe Gly Asp Phe Val Gln Val Ala Pro Gly Cys Gly Val Pro Ile
 210 215 220
 Val Asp Ser Tyr Tyr Ser Leu Leu Met Pro Ile Leu Thr Leu Thr Arg
 225 230 235 240
 Ala Leu Ala Ala Glu Ser His Met Asp Ala Asp Leu Ala Lys Pro Leu
 245 250 255
 Ile Lys Trp Asp Leu Leu Lys Tyr Asp Phe Thr Glu Glu Arg Leu Cys
 260 265 270

S226CAS108.ST25

Leu Phe Asp Arg Tyr Phe Lys Tyr Trp Asp Gln Thr Tyr His Pro Asn
 275 280 285
 Cys Ile Asn Cys Leu Asp Asp Arg Cys Ile Leu His Cys Ala Asn Phe
 290 295 300
 Asn Val Leu Phe Ser Thr Val Phe Pro Pro Thr Ser Phe Gly Pro Leu
 305 310 315 320
 Val Arg Lys Ile Phe Val Asp Gly Val Pro Phe Val Val Ser Thr Gly
 325 330 335
 Tyr His Phe Arg Glu Leu Gly Val Val His Asn Gln Asp Val Asn Leu
 340 345 350
 His Ser Ser Arg Leu Ser Phe Lys Glu Leu Leu Val Tyr Ala Ala Asp
 355 360 365
 Pro Ala Met His Ala Ala Ser Gly Asn Leu Leu Leu Asp Lys Arg Thr
 370 375 380
 Thr Cys Phe Ser Val Ala Ala Leu Thr Asn Asn Val Ala Phe Gln Thr
 385 390 395 400
 Val Lys Pro Gly Asn Phe Asn Lys Asp Phe Tyr Asp Phe Ala Val Ser
 405 410 415
 Lys Gly Phe Phe Lys Glu Gly Ser Ser Val Glu Leu Lys His Phe Phe
 420 425 430
 Phe Ala Gln Asp Gly Asn Ala Ala Ile Ser Asp Tyr Asp Tyr Tyr Arg
 435 440 445
 Tyr Asn Leu Pro Thr Met Cys Asp Ile Arg Gln Leu Leu Phe Val Val
 450 455 460
 Glu Val Val Asp Lys Tyr Phe Asp Cys Tyr Asp Gly Gly Cys Ile Asn
 465 470 475 480
 Ala Asn Gln Val Ile Val Asn Asn Leu Asp Lys Ser Ala Gly Phe Pro
 485 490 495
 Phe Asn Lys Trp Gly Lys Ala Arg Leu Tyr Tyr Asp Ser Met Ser Tyr
 500 505 510
 Glu Asp Gln Asp Ala Leu Phe Ala Tyr Thr Lys Arg Asn Val Ile Pro
 515 520 525
 Thr Ile Thr Gln Met Asn Leu Lys Tyr Ala Ile Ser Ala Lys Asn Arg
 530 535 540

S226CAS108.ST25

Ala Arg Thr Val Ala Gly Val Ser Ile Cys Ser Thr Met Thr Asn Arg
 545 550 555 560
 Gln Phe His Gln Lys Leu Leu Lys Ser Ile Ala Ala Thr Arg Gly Ala
 565 570 575
 Thr Val Val Ile Gly Thr Ser Lys Phe Tyr Gly Gly Trp His Asn Met
 580 585 590
 Leu Lys Thr Val Tyr Ser Asp Val Glu Thr Pro His Leu Met Gly Trp
 595 600 605
 Asp Tyr Pro Lys Cys Asp Arg Ala Met Pro Asn Met Leu Arg Ile Met
 610 615 620
 Ala Ser Leu Val Leu Ala Arg Lys His Asn Thr Cys Cys Asn Leu Ser
 625 630 635 640
 His Arg Phe Tyr Arg Leu Ala Asn Glu Cys Ala Gln Val Leu Ser Glu
 645 650 655
 Met Val Met Cys Gly Gly Ser Leu Tyr Val Lys Pro Gly Gly Thr Ser
 660 665 670
 Ser Gly Asp Ala Thr Thr Ala Tyr Ala Asn Ser Val Phe Asn Ile Cys
 675 680 685
 Gln Ala Val Thr Ala Asn Val Asn Ala Leu Leu Ser Thr Asp Gly Asn
 690 695 700
 Lys Ile Ala Asp Lys Tyr Val Arg Asn Leu Gln His Arg Leu Tyr Glu
 705 710 715 720
 Cys Leu Tyr Arg Asn Arg Asp Val Asp His Glu Phe Val Asp Glu Phe
 725 730 735
 Tyr Ala Tyr Leu Arg Lys His Phe Ser Met Met Ile Leu Ser Asp Asp
 740 745 750
 Ala Val Val Cys Tyr Asn Ser Asn Tyr Ala Ala Gln Gly Leu Val Ala
 755 760 765
 Ser Ile Lys Asn Phe Lys Ala Val Leu Tyr Tyr Gln Asn Asn Val Phe
 770 775 780
 Met Ser Glu Ala Lys Cys Trp Thr Glu Thr Asp Leu Thr Lys Gly Pro
 785 790 795 800
 His Glu Phe Cys Ser Gln His Thr Met Leu Val Lys Gln Gly Asp Asp
 805 810 815

S226CAS108.ST25

Tyr Val Tyr Leu Pro Tyr Pro Asp Pro Ser Arg Ile Leu Gly Ala Gly
 820 825 830
 Cys Phe Val Asp Asp Ile Val Lys Thr Asp Gly Thr Leu Met Ile Glu
 835 840 845
 Arg Phe Val Ser Leu Ala Ile Asp Ala Tyr Pro Leu Thr Lys His Pro
 850 855 860
 Asn Gln Glu Tyr Ala Asp Val Phe His Leu Tyr Leu Gln Tyr Ile Arg
 865 870 875 880
 Lys Leu His Asp Glu Leu Thr Gly His Met Leu Asp Met Tyr Ser Val
 885 890 895
 Met Leu Thr Asn Asp Asn Thr Ser Arg Tyr Trp Glu Pro Glu Phe Tyr
 900 905 910
 Glu Ala Met Tyr Thr Pro His Thr Val Leu Gln Ala Val Gly Ala Cys
 915 920 925
 Val Leu Cys Asn Ser Gln Thr Ser Leu Arg Cys Gly Ala Cys Ile Arg
 930 935 940
 Arg Pro Phe Leu Cys Cys Lys Cys Cys Tyr Asp His Val Ile Ser Thr
 945 950 955 960
 Ser His Lys Leu Val Leu Ser Val Asn Pro Tyr Val Cys Asn Ala Pro
 965 970 975
 Gly Cys Asp Val Thr Asp Val Thr Gln Leu Tyr Leu Gly Gly Met Ser
 980 985 990
 Tyr Tyr Cys Lys Ser His Lys Pro Pro Ile Ser Phe Pro Leu Cys Ala
 995 1000 1005
 Asn Gly Gln Val Phe Gly Leu Tyr Lys Asn Thr Cys Val Gly Ser
 1010 1015 1020
 Asp Asn Val Thr Asp Phe Asn Ala Ile Ala Thr Cys Asp Trp Thr
 1025 1030 1035
 Asn Ala Gly Asp Tyr Ile Leu Ala Asn Thr Cys Thr Glu Arg Leu
 1040 1045 1050
 Lys Leu Phe Ala Ala Glu Thr Leu Lys Ala Thr Glu Glu Thr Phe
 1055 1060 1065
 Lys Leu Ser Tyr Gly Ile Ala Thr Val Arg Glu Val Leu Ser Asp
 1070 1075 1080

S226CAS108.ST25

Arg	Glu 1085	Leu	His	Leu	Ser	Trp 1090	Glu	Val	Gly	Lys	Pro 1095	Arg	Pro	Pro
Leu	Asn 1100	Arg	Asn	Tyr	Val	Phe 1105	Thr	Gly	Tyr	Arg	Val 1110	Thr	Lys	Asn
Ser	Lys 1115	Val	Gln	Ile	Gly	Glu 1120	Tyr	Thr	Phe	Glu	Lys 1125	Gly	Asp	Tyr
Gly	Asp 1130	Ala	Val	Val	Tyr	Arg 1135	Gly	Thr	Thr	Thr	Tyr 1140	Lys	Leu	Asn
Val	Gly 1145	Asp	Tyr	Phe	Val	Leu 1150	Thr	Ser	His	Thr	Val 1155	Met	Pro	Leu
Ser	Ala 1160	Pro	Thr	Leu	Val	Pro 1165	Gln	Glu	His	Tyr	Val 1170	Arg	Ile	Thr
Gly	Leu 1175	Tyr	Pro	Thr	Leu	Asn 1180	Ile	Ser	Asp	Glu	Phe 1185	Ser	Ser	Asn
Val	Ala 1190	Asn	Tyr	Gln	Lys	Val 1195	Gly	Met	Gln	Lys	Tyr 1200	Ser	Thr	Leu
Gln	Gly 1205	Pro	Pro	Gly	Thr	Gly 1210	Lys	Ser	His	Phe	Ala 1215	Ile	Gly	Leu
Ala	Leu 1220	Tyr	Tyr	Pro	Ser	Ala 1225	Arg	Ile	Val	Tyr	Thr 1230	Ala	Cys	Ser
His	Ala 1235	Ala	Val	Asp	Ala	Leu 1240	Cys	Glu	Lys	Ala	Leu 1245	Lys	Tyr	Leu
Pro	Ile 1250	Asp	Lys	Cys	Ser	Arg 1255	Ile	Ile	Pro	Ala	Arg 1260	Ala	Arg	Val
Glu	Cys 1265	Phe	Asp	Lys	Phe	Lys 1270	Val	Asn	Ser	Thr	Leu 1275	Glu	Gln	Tyr
Val	Phe 1280	Cys	Thr	Val	Asn	Ala 1285	Leu	Pro	Glu	Thr	Thr 1290	Ala	Asp	Ile
Val	Val 1295	Phe	Asp	Glu	Ile	Ser 1300	Met	Ala	Thr	Asn	Tyr 1305	Asp	Leu	Ser
Val	Val 1310	Asn	Ala	Arg	Leu	Arg 1315	Ala	Lys	His	Tyr	Val 1320	Tyr	Ile	Gly
Asp	Pro 1325	Ala	Gln	Leu	Pro	Ala 1330	Pro	Arg	Thr	Leu	Leu 1335	Thr	Lys	Gly

S226CAS108.ST25

Thr Leu Glu Pro Glu Tyr Phe Asn Ser Val Cys Arg Leu Met Lys
 1340 1345 1350
 Thr Ile Gly Pro Asp Met Phe Leu Gly Thr Cys Arg Arg Cys Pro
 1355 1360 1365
 Ala Glu Ile Val Asp Thr Val Ser Ala Leu Val Tyr Asp Asn Lys
 1370 1375 1380
 Leu Lys Ala His Lys Asp Lys Ser Ala Gln Cys Phe Lys Met Phe
 1385 1390 1395
 Tyr Lys Gly Val Ile Thr His Asp Val Ser Ser Ala Ile Asn Arg
 1400 1405 1410
 Pro Gln Ile Gly Val Val Arg Glu Phe Leu Thr Arg Asn Pro Ala
 1415 1420 1425
 Trp Arg Lys Ala Val Phe Ile Ser Pro Tyr Asn Ser Gln Asn Ala
 1430 1435 1440
 Val Ala Ser Lys Ile Leu Gly Leu Pro Thr Gln Thr Val Asp Ser
 1445 1450 1455
 Ser Gln Gly Ser Glu Tyr Asp Tyr Val Ile Phe Thr Gln Thr Thr
 1460 1465 1470
 Glu Thr Ala His Ser Cys Asn Val Asn Arg Phe Asn Val Ala Ile
 1475 1480 1485
 Thr Arg Ala Lys Ile Gly Ile Leu Cys Ile Met Ser Asp Arg Asp
 1490 1495 1500
 Leu Tyr Asp Lys Leu Gln Phe Thr Ser Leu Glu Ile Pro Arg Arg
 1505 1510 1515
 Asn Val Ala Thr Leu Gln Ala Glu Asn Val Thr Gly Leu Phe Lys
 1520 1525 1530
 Asp Cys Ser Lys Ile Ile Thr Gly Leu His Pro Thr Gln Ala Pro
 1535 1540 1545
 Thr His Leu Ser Val Asp Ile Lys Phe Lys Thr Glu Gly Leu Cys
 1550 1555 1560
 Val Asp Ile Pro Gly Ile Pro Lys Asp Met Thr Tyr Arg Arg Leu
 1565 1570 1575
 Ile Ser Met Met Gly Phe Lys Met Asn Tyr Gln Val Asn Gly Tyr
 1580 1585 1590

S226CAS108.ST25

Pro Asn Met Phe Ile Thr Arg Glu Glu Ala Ile Arg His Val Arg
 1595 1600 1605
 Ala Trp Ile Gly Phe Asp Val Glu Gly Cys His Ala Thr Arg Asp
 1610 1615 1620
 Ala Val Gly Thr Asn Leu Pro Leu Gln Leu Gly Phe Ser Thr Gly
 1625 1630 1635
 Val Asn Leu Val Ala Val Pro Thr Gly Tyr Val Asp Thr Glu Asn
 1640 1645 1650
 Asn Thr Glu Phe Thr Arg Val Asn Ala Lys Pro Pro Pro Gly Asp
 1655 1660 1665
 Gln Phe Lys His Leu Ile Pro Leu Met Tyr Lys Gly Leu Pro Trp
 1670 1675 1680
 Asn Val Val Arg Ile Lys Ile Val Gln Met Leu Ser Asp Thr Leu
 1685 1690 1695
 Lys Gly Leu Ser Asp Arg Val Val Phe Val Leu Trp Ala His Gly
 1700 1705 1710
 Phe Glu Leu Thr Ser Met Lys Tyr Phe Val Lys Ile Gly Pro Glu
 1715 1720 1725
 Arg Thr Cys Cys Leu Cys Asp Lys Arg Ala Thr Cys Phe Ser Thr
 1730 1735 1740
 Ser Ser Asp Thr Tyr Ala Cys Trp Asn His Ser Val Gly Phe Asp
 1745 1750 1755
 Tyr Val Tyr Asn Pro Phe Met Ile Asp Val Gln Gln Trp Gly Phe
 1760 1765 1770
 Thr Gly Asn Leu Gln Ser Asn His Asp Gln His Cys Gln Val His
 1775 1780 1785
 Gly Asn Ala His Val Ala Ser Cys Asp Ala Ile Met Thr Arg Cys
 1790 1795 1800
 Leu Ala Val His Glu Cys Phe Val Lys Arg Val Asp Trp Ser Val
 1805 1810 1815
 Glu Tyr Pro Ile Ile Gly Asp Glu Leu Arg Val Asn Ser Ala Cys
 1820 1825 1830
 Arg Lys Val Gln His Met Val Val Lys Ser Ala Leu Leu Ala Asp
 1835 1840 1845

S226CAS108.ST25

Lys Phe Pro Val Leu His Asp Ile Gly Asn Pro Lys Ala Ile Lys
 1850 1855 1860
 Cys Val Pro Gln Ala Glu Val Glu Trp Lys Phe Tyr Asp Ala Gln
 1865 1870 1875
 Pro Cys Ser Asp Lys Ala Tyr Lys Ile Glu Glu Leu Phe Tyr Ser
 1880 1885 1890
 Tyr Ala Thr His His Asp Lys Phe Thr Asp Gly Val Cys Leu Phe
 1895 1900 1905
 Trp Asn Cys Asn Val Asp Arg Tyr Pro Ala Asn Ala Ile Val Cys
 1910 1915 1920
 Arg Phe Asp Thr Arg Val Leu Ser Asn Leu Asn Leu Pro Gly Cys
 1925 1930 1935
 Asp Gly Gly Ser Leu Tyr Val Asn Lys His Ala Phe His Thr Pro
 1940 1945 1950
 Ala Phe Asp Lys Ser Ala Phe Thr Asn Leu Lys Gln Leu Pro Phe
 1955 1960 1965
 Phe Tyr Tyr Ser Asp Ser Pro Cys Glu Ser His Gly Lys Gln Val
 1970 1975 1980
 Val Ser Asp Ile Asp Tyr Val Pro Leu Lys Ser Ala Thr Cys Ile
 1985 1990 1995
 Thr Arg Cys Asn Leu Gly Gly Ala Val Cys Arg His His Ala Asn
 2000 2005 2010
 Glu Tyr Arg Gln Tyr Leu Asp Ala Tyr Asn Met Met Ile Ser Ala
 2015 2020 2025
 Gly Phe Ser Leu Trp Ile Tyr Lys Gln Phe Asp Thr Tyr Asn Leu
 2030 2035 2040
 Trp Asn Thr Phe Thr Arg Leu Gln Ser Leu Glu Asn Val Ala Tyr
 2045 2050 2055
 Asn Val Val Asn Lys Gly His Phe Asp Gly His Ala Gly Glu Ala
 2060 2065 2070
 Pro Val Ser Ile Ile Asn Asn Ala Val Tyr Thr Lys Val Asp Gly
 2075 2080 2085
 Ile Asp Val Glu Ile Phe Glu Asn Lys Thr Thr Leu Pro Val Asn
 2090 2095 2100

S226CAS108.ST25

Val Ala Phe Glu Leu Trp Ala Lys Arg Asn Ile Lys Pro Val Pro
 2105 2110 2115
 Glu Ile Lys Ile Leu Asn Asn Leu Gly Val Asp Ile Ala Ala Asn
 2120 2125 2130
 Thr Val Ile Trp Asp Tyr Lys Arg Glu Ala Pro Ala His Val Ser
 2135 2140 2145
 Thr Ile Gly Val Cys Thr Met Thr Asp Ile Ala Lys Lys Pro Thr
 2150 2155 2160
 Glu Ser Ala Cys Ser Ser Leu Thr Val Leu Phe Asp Gly Arg Val
 2165 2170 2175
 Glu Gly Gln Val Asp Leu Phe Arg Asn Ala Arg Asn Gly Val Leu
 2180 2185 2190
 Ile Thr Glu Gly Ser Val Lys Gly Leu Thr Pro Ser Lys Gly Pro
 2195 2200 2205
 Ala Gln Ala Ser Val Asn Gly Val Thr Leu Ile Gly Glu Ser Val
 2210 2215 2220
 Lys Thr Gln Phe Asn Tyr Phe Lys Lys Val Asp Gly Ile Ile Gln
 2225 2230 2235
 Gln Leu Pro Glu Thr Tyr Phe Thr Gln Ser Arg Asp Leu Glu Asp
 2240 2245 2250
 Phe Lys Pro Arg Ser Gln Met Glu Thr Asp Phe Leu Glu Leu Ala
 2255 2260 2265
 Met Asp Glu Phe Ile Gln Arg Tyr Lys Leu Glu Gly Tyr Ala Phe
 2270 2275 2280
 Glu His Ile Val Tyr Gly Asp Phe Ser His Gly Gln Leu Gly Gly
 2285 2290 2295
 Leu His Leu Met Ile Gly Leu Ala Lys Arg Ser Gln Asp Ser Pro
 2300 2305 2310
 Leu Lys Leu Glu Asp Phe Ile Pro Met Asp Ser Thr Val Lys Asn
 2315 2320 2325
 Tyr Phe Ile Thr Asp Ala Gln Thr Gly Ser Ser Lys Cys Val Cys
 2330 2335 2340
 Ser Val Ile Asp Leu Leu Leu Asp Asp Phe Val Glu Ile Ile Lys
 2345 2350 2355

S226CAS108.ST25

Ser Gln Asp Leu Ser Val Ile Ser Lys Val Val Lys Val Thr Ile
 2360 2365 2370
 Asp Tyr Ala Glu Ile Ser Phe Met Leu Trp Cys Lys Asp Gly His
 2375 2380 2385
 Val Glu Thr Phe Tyr Pro Lys Leu Gln Ala Ser Gln Ala Trp Gln
 2390 2395 2400
 Pro Gly Val Ala Met Pro Asn Leu Tyr Lys Met Gln Arg Met Leu
 2405 2410 2415
 Leu Glu Lys Cys Asp Leu Gln Asn Tyr Gly Glu Asn Ala Val Ile
 2420 2425 2430
 Pro Lys Gly Ile Met Met Asn Val Ala Lys Tyr Thr Gln Leu Cys
 2435 2440 2445
 Gln Tyr Leu Asn Thr Leu Thr Leu Ala Val Pro Tyr Asn Met Arg
 2450 2455 2460
 Val Ile His Phe Gly Ala Gly Ser Asp Lys Gly Val Ala Pro Gly
 2465 2470 2475
 Thr Ala Val Leu Arg Gln Trp Leu Pro Thr Gly Thr Leu Leu Val
 2480 2485 2490
 Asp Ser Asp Leu Asn Asp Phe Val Ser Asp Ala Asp Ser Thr Leu
 2495 2500 2505
 Ile Gly Asp Cys Ala Thr Val His Thr Ala Asn Lys Trp Asp Leu
 2510 2515 2520
 Ile Ile Ser Asp Met Tyr Asp Pro Arg Thr Lys His Val Thr Lys
 2525 2530 2535
 Glu Asn Asp Ser Lys Glu Gly Phe Phe Thr Tyr Leu Cys Gly Phe
 2540 2545 2550
 Ile Lys Gln Lys Leu Ala Leu Gly Gly Ser Ile Ala Val Lys Ile
 2555 2560 2565
 Thr Glu His Ser Trp Asn Ala Asp Leu Tyr Lys Leu Met Gly His
 2570 2575 2580
 Phe Ser Trp Trp Thr Ala Phe Val Thr Asn Val Asn Ala Ser Ser
 2585 2590 2595
 Ser Glu Ala Phe Leu Ile Gly Ala Asn Tyr Leu Gly Lys Pro Lys
 2600 2605 2610

S226CAS108.ST25

Glu Gln Ile Asp Gly Tyr Thr Met His Ala Asn Tyr Ile Phe Trp
 2615 2620 2625

Arg Asn Thr Asn Pro Ile Gln Leu Ser Ser Tyr Ser Leu Phe Asp
 2630 2635 2640

Met Ser Lys Phe Pro Leu Lys Leu Arg Gly Thr Ala Val Met Ser
 2645 2650 2655

Leu Lys Glu Asn Gln Ile Asn Asp Met Ile Tyr Ser Leu Leu Glu
 2660 2665 2670

Lys Gly Arg Leu Ile Ile Arg Glu Asn Asn Arg Val Val Val Ser
 2675 2680 2685

Ser Asp Ile Leu Val Asn Asn
 2690 2695

<210> 76

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L3/+/4932

<400> 76
 ccacacacag cttgtggata

20

<210> 77

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L4/+/6401

<400> 77
 ccgaagttgt aggcaatgtc

20

<210> 78

<211> 20

<212> DNA

S226CAS108.ST25

<213> Artificial sequence

<220>

<223> amorce S/L4/+/6964

<400> 78

tttggtgctc cttcttattg

20

<210> 79

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L4/-/6817

<400> 79

ccggcatcca aacataattt

20

<210> 80

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L5/-/7633

<400> 80

tggtcagtag gggtgattgg

20

<210> 81

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> S/L5/-/8127

<400> 81

catcctttgt gtcaacatcg

20

<210> 82

S226CAS108.ST25

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L5/-/8633

<400> 82

gtcacgagtg acaccatcct

20

<210> 83

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L5+/7839

<400> 83

atgcgacgag tctgcttcta

20

<210> 84

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L5+/8785

<400> 84

ttcatagtgctg ctggcttacc

20

<210> 85

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L5+/8255

<400> 85

atcttggcgc atgtattgac

20

S226CAS108.ST25

<210> 86
<211> 20
<212> DNA
<213> Artificial sequence

<220>

<223> amorce S/L6/-/9422

<400> 86
tgcattagca gcaacaacat 20

<210> 87
<211> 20
<212> DNA
<213> Artificial sequence

<220>

<223> amorce S/L6/-/9966

<400> 87
tctgcagaac agcagaagtg 20

<210> 88
<211> 20
<212> DNA
<213> Artificial sequence

<220>

<223> amorce S/L6/-/10542

<400> 88
cctgtgcagt ttgtctgtca 20

<210> 89
<211> 20
<212> DNA
<213> Artificial sequence

<220>

S226CAS108.ST25

<223> amorce S/L6/+ /10677

<400> 89

ccttggtggca atgaagtaca

20

<210> 90

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L6/+ /10106

<400> 90

atgtcatttg cacagcagaa

20

<210> 91

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L6/+ /9571

<400> 91

cttcaatggg ttgccatggt

20

<210> 92

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L7/- /11271

<400> 92

tgcgagctgt catgagaata

20

<210> 93

<211> 20

<212> DNA

<213> Artificial sequence

S226CAS108.ST25

<220>

<223> amorce S/L7/-/11801

<400> 93

aaccgagagc agtaccacag

20

<210> 94

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L7/-/12383

<400> 94

tttggctgct gtagtcaatg

20

<210> 95

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L7+/12640

<400> 95

ctacgacaga tgcctgtgc

20

<210> 96

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L7+/12088

<400> 96

gagcaggctg tagctaattg

20

<210> 97

<211> 20

S226CAS108.ST25

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L7/+/11551

<400> 97

ttaggctatt gttgctgctg

20

<210> 98

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L8/-/13160

<400> 98

cagacaacat gaagcaccac

20

<210> 99

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L8/-/13704

<400> 99

cgctgacgtg atatatgtgg

20

<210> 100

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L8/-/14284

<400> 100

tgcacaatga aggatacacc

20

S226CAS108.ST25

<210> 101
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> amorce S/L8/+/14453

<400> 101
acatagctcg cgtctcagtt 20

<210> 102
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> amorce S/L8/+/13968

<400> 102
ggcattgtag gcgtactgac 20

<210> 103
<211> 19
<212> DNA
<213> Artificial sequence

<220>
<223> amorce S/L8/+/13401

<400> 103
gtttgcggtg taagtgcag 19

<210> 104
<211> 20
<212> DNA
<213> Artificial sequence

<220>
<223> amorce S/L9/-/15098

S226CAS108.ST25

<400> 104
tagtggcggc tattgacttc

20

<210> 105

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L9/-/15677

<400> 105
ctaaaccttg agccgcatag

20

<210> 106

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L9/-/16247

<400> 106
catggtcata gcagcacttg

20

<210> 107

<211> 21

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L9/+/16323

<400> 107
ccaggttgtg atgtcactga t

21

<210> 108

<211> 20

<212> DNA

<213> Artificial sequence

S226CAS108.ST25

<220>

<223> amorce S/L9/+/15858

<400> 108

ccttaccag atccatcaag

20

<210> 109

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L9/+/15288

<400> 109

cgcaaacata acacttgctg

20

<210> 110

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L10/-/16914

<400> 110

agtgttggt acaagccagt

20

<210> 111

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L10/-/17466

<400> 111

gttccaagga acatgtctgg

20

<210> 112

<211> 20

<212> DNA

S226CAS108.ST25

<213> Artificial sequence

<220>

<223> amorce S/L10/-/18022

<400> 112

aggtgcctgt gtaggatgaa

20

<210> 113

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L10/+18245

<400> 113

gggctgtcat gcaactagag

20

<210> 114

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L10/+17663

<400> 114

tcttacacgc aatcctgctt

20

<210> 115

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L10/+17061

<400> 115

taccatctg ctcgcatagt

20

<210> 116

S226CAS108.ST25

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L11/-/18877

<400> 116

gcaagcagaa ttaaccctca

20

<210> 117

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L11/-/19396

<400> 117

agcaccacct aaattgcatc

20

<210> 118

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L11/-/20002

<400> 118

tggtcccttt gaaggtgtta

20

<210> 119

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L11/+ /20245

<400> 119

tcgaacacat cgtttatgga

20

S226CAS108.ST25

<210> 120

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L11/+/19611

<400> 120

gaagcacctg tttccatcat

20

<210> 121

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce S/L11/+/19021

<400> 121

acgatgctca gccatgtagt

20

<210> 122

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L1/F3/+/800

<400> 122

gaggtgcagt cactcgctat

20

<210> 123

<211> 20

<212> DNA

<213> Artificial sequence

<220>

S226CAS108.ST25

<223> amorce SARS/L1/F4/+/1391

<400> 123

cagagattgg acctgagcat

20

<210> 124

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L1/F5/+/1925

<400> 124

cagcaaacca ctcaattcct

20

<210> 125

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L1/R3/-/1674

<400> 125

aaatgatggc aacctcttca

20

<210> 126

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L1/R4/-/1107

<400> 126

cacgtggttg aatgactttg

20

<210> 127

<211> 20

<212> DNA

<213> Artificial sequence

S226CAS108.ST25

<220>

<223> amorce SARS/L1/R5/-/520

<400> 127

atttctgcaa ccagctcaac

20

<210> 128

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L2/F3/+ /2664

<400> 128

cgcattgtct cctggtttac

20

<210> 129

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L2/F4/+ /3232

<400> 129

gagattgagc cagaaccaga

20

<210> 130

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L2/F5/+ /3746

<400> 130

atgagcaggt tgtcatggat

20

<210> 131

<211> 20

S226CAS108.ST25

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L2/R3/-/3579

<400> 131

ctgccttaag aagctggatg

20

<210> 132

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L2/R4/-/2991

<400> 132

tttcttcacc agcatcatca

20

<210> 133

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L2/R5/-/2529

<400> 133

caccgttctt gagaacaacc

20

<210> 134

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L3/F3/+/4708

<400> 134

tctttggctg gctcttacag

20

S226CAS108.ST25

<210> 135

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SRAS/L3/F4/+ /5305

<400> 135

gctggatgatg ctgctaactt

20

<210> 136

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L3/F5/+ /5822

<400> 136

ccatcaagcc tgtgtcgtat

20

<210> 137

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L3/R3/- /5610

<400> 137

caggtggtgc agacatcata

20

<210> 138

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L3/R4/- /4988

S226CAS108.ST25

<400> 138
aacatcagca ccatccaagt

20

<210> 139

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SARS/L3/R5/-/4437

<400> 139
atcggacacc atagtcaacg

20

<210> 140

<211> 7788

<212> DNA

<213> Artificial sequence

<220>

<223> gène S synthétique

<400> 140
tcaatattgg ccattagcca tattattcat tggttatata gcataaatca atattggcta 60
ttggccattg catacgttgt atctatatca taatatgtac atttatattg gtcattgtcc 120
aatatgaccg ccatgttggc attgattatt gactagttat taatagtaat caattacggg 180
gtcattagtt catagcccat atatggagtt ccgcgttaca taacttacgg taaatggccc 240
gcctggctga ccgccaacg acccccgcgc attgacgtca ataatgacgt atgttcccat 300
agtaacgcca atagggactt tccattgacg tcaatgggtg gagtatttac ggtaaactgc 360
ccacttggca gtacatcaag tgtatcatat gccaaagtccg cccctattg acgtcaatga 420
cggtaaatgg cccgcctggc attatgcca gtacatgacc ttacgggact ttcctacttg 480
gcagtacatc tacgtattag tcatcgctat taccatgggtg atgcgggtttt ggcagtacac 540
caatgggctg ggatagcggg ttgactcacg gggatttcca agtctccacc ccattgacgt 600
caatgggagt ttgttttggc accaaaatca acgggacttt ccaaaatgtc gtaataaccc 660
cgccccgttg acgcaaattg gcggtaggcg tgtacgggtg gaggtctata taagcagagc 720
tcgttttagtg aaccgtcaga tactagaag ctttattgctg gtagtttatc acagttaaatt 780
tgctaacgca gtcagtgttt ctgacacaac agtctcgaac ttaagctgca gaagttggctc 840
gtgaggcact gggcaggtaa gtatcaagggt tacaagacag gtttaaggag accaatagaa 900
actgggcttg tcgagacaga gaagactctt gcgttttctga taggcaccta ttggtcttac 960

S226CAS108.ST25

tgacatccac	tttgcctttc	tctccacagg	tgtccactcc	cagttcaatt	acagctctta	1020
aggctagagt	acttaatacg	actcactata	ggctagcgga	tccaccatgt	tcattcttct	1080
gctgttcctg	accctgacca	gcggcagcga	cctggaccgg	tgcaccacct	tcgacgacgt	1140
gcaggccccc	aactacaccc	agcacaccag	cagcatgcgg	ggcgtgtact	accccagcga	1200
gatctttcgg	agcgacaccc	tgtacctgac	ccaggacctg	ttcctgcccct	tctacagcaa	1260
cgtgaccggc	ttccacacca	tcaaccacac	cttcggcaac	cccgtgatcc	ccttcaagga	1320
cggcatctac	ttcgccgcca	ccgagaagag	caacgtggtg	cggggctggg	tgttcggcag	1380
caccatgaac	aacaagagcc	agagcgtgat	catcatcaac	aacagcacca	acgtggtgat	1440
ccgggcctgc	aacttcgagc	tgtgcgacaa	ccccttcttc	gccgtgtcca	aacctatggg	1500
cacccagacc	cacaccatga	tcttcgacaa	cgccctcaac	tgcaccttcg	agtacatcag	1560
cgacgccttc	agcctggacg	tgagcgagaa	gagcggcaac	ttcaagcacc	tgcgggagtt	1620
cgtgttcaag	aacaaggacg	gcttcctgta	cgtgtacaag	ggctaccagc	ccatcgacgt	1680
ggtgagagac	ctgcccagcg	gcttcaacac	cctgaagccc	atcttcaagc	tgcccctggg	1740
catcaacatc	accaacttcc	gggccatcct	gaccgccttt	agccctgccc	aggacatctg	1800
gggcaccagc	gccgccgcct	acttcgtggg	ctacctgaag	cctaccacct	tcattgctgaa	1860
gtacgacgag	aacggcacca	tcaccgacgc	cgtggactgc	agccagaacc	ccctggccga	1920
gctgaagtgc	agcgtgaaga	gcttcgagat	cgacaagggc	atctaccaga	ccagcaactt	1980
cagagtgggtg	cctagcggcg	atgtgggtcg	gttccccaat	atcaccaacc	tgtgcccctt	2040
cggcgaagtg	ttcaacgcca	ccaagttccc	cagcgtgtac	gcctgggagc	ggaagaagat	2100
cagcaactgc	gtggccgact	acagcgtgct	gtacaactcc	accttcttca	gcaccttcaa	2160
gtgctacggc	gtgagcgcca	ccaagctgaa	cgacctgtgc	ttcagcaacg	tgtacgccga	2220
cagcttcgtg	gtgaagggcg	acgacgtgag	acagatcgcc	cctggccaga	ccggcgtgat	2280
cgccgactac	aactacaagc	tgcccagacga	cttcattggc	tgctgtctgg	cctggaacac	2340
ccggaacatc	gacgccacaa	gcaccggcaa	ctacaattac	aagtaccgct	acctgcggca	2400
cggcaagtgc	cggcccttcg	agcgggacat	ctccaacgtg	cccttcagcc	ccgacggcaa	2460
gccctgcacc	ccccctgccc	tgaactgcta	ctggcccctg	aacgactacg	gcttctacac	2520
caccaccggc	atcggtatc	agccctacag	agtgggtggtg	ctgagcttcg	agctgctgaa	2580
cggccctgcc	accgtgtgcg	gccccaaagct	gagcaccgac	ctgatcaaga	accagtgcgt	2640
gaacttcaac	ttcaacggcc	tgaccggcac	cggcgtgctg	acccccagca	gcaagcgctt	2700
ccagcccttc	cagcagttcg	gccgggatgt	gagcgacttc	accgacagcg	tgcgggaccc	2760
caagaccagc	gagatcctgg	acatcagccc	ctgcagcttc	ggcggcgtgt	ccgtgatcac	2820
ccccggcacc	aacgccagca	gcgaagtggc	cgtgctgtac	caggacgtga	actgcaccga	2880
cgtgagcacc	gccatccacg	ccgaccagct	gacccccgcc	tggcggatct	acagcaccgg	2940
gaacaacgtg	ttccagaccc	aggccggctg	cctgatcggc	gccgagcacg	tggacaccag	3000

S226CAS108.ST25

ctacgagtg	gacatcccca	ttggcgccgg	aatctgcgcc	agctaccaca	ccgtgagcct	3060
gctgcggagc	accagccaga	agtccatcgt	ggcctacacc	atgagcctgg	gcgccgacag	3120
cagcatcgcc	tacagcaaca	acaccatcgc	catccccacc	aacttcagca	tctccatcac	3180
caccgaagt	atgcccgta	gcatggccaa	gacaagcgtg	gattgcaaca	tgtacatctg	3240
cggcgacagc	accgagtgcg	ccaacctgct	gctgcagtac	ggcagcttct	gcacccagct	3300
gaaccggggc	ctgagcggca	tcgccgccga	gcaggaccgg	aacaccagag	aagtgttcgc	3360
ccaagtgaag	cagatgtata	agacccccac	cctgaagtac	ttcgggggct	tcaacttctc	3420
tcagatcctg	cccgaccctc	tgaagcccac	caagcgctcc	ttcatcgagg	acctgctggt	3480
caacaaagt	accctggccg	acgccggctt	tatgaagcag	tacggcgagt	gcctgggcga	3540
catcaacgcc	cgggacctga	tctgcgcccc	gaagtttaac	gggctgaccg	tgctgcccc	3600
cctgctgacc	gacgacatga	tcgccgccta	tacagccgcc	ctggtgagcg	gcaccgccac	3660
cgccggctgg	accttcggag	ccggagccgc	cctgcagatc	cccttcgcc	tgagatggc	3720
ctaccggttc	aacggcatcg	gcgtgaccca	gaacgtgctg	tacgagaacc	agaagcagat	3780
cgccaaccag	ttcaacaagg	ccatcagcca	gatccaggag	agcctgacca	caaccagcac	3840
cggcctgggc	aagctgcagg	acgtggtgaa	ccagaacgcc	caggccctga	acaccctggt	3900
gaagcagctg	agcagcaact	tcggcgccat	cagctctgtg	ctgaacgaca	tcctgagcag	3960
gctggacaaa	gtggaggccg	aagtgcagat	cgaccggctg	atcaccggac	gcctgcagtc	4020
cctgcagacc	tacgtgacct	agcagctgat	cagagccgcc	gagatccggg	ccagcgccaa	4080
tctggccgcc	accaagatga	gcgagtgcgt	gctgggccag	agcaagagag	tggacttctg	4140
cggcaagggc	tatcacctga	tgagcttccc	ccaggccgcc	ccccacggcg	tggtgttctt	4200
gcacgtgacc	tacgtgccta	gccaggagcg	gaacttcacc	accgccccag	ccatctgcc	4260
cgagggcaag	gcctacttcc	ccggggaggg	cgtgttcgtg	tttaacggca	ccagctgggt	4320
catcaccag	cgcaacttct	tcagccccc	gatcatcacc	acagacaaca	ccttcgtgtc	4380
cggcaactgt	gatgtggtga	tcggcatcat	caataacacc	gtgtacgacc	ccctgcagcc	4440
cgagctggac	agcttcaagg	aggagctgga	caaatacttc	aagaaccaca	cctccccga	4500
cgtggacctg	ggcgatatca	gcggcatcaa	cgctccgtg	gtgaacatcc	agaaggagat	4560
cgacagactg	aacgaagtgg	ccaagaacct	gaacgagagc	ctgatcgacc	tgaggagct	4620
gggcaagtac	gagcagtaca	tcaagtggcc	ctggtacgtg	tggctgggct	tcatcgccgg	4680
cctgatcgcc	atcgtgatgg	tgaccatcct	gctgtgctgc	atgaccagct	gctgtagctg	4740
cctgaaaggc	gcctgcagct	gtggcagctg	ctgcaagttc	gacgaggacg	acagcgagcc	4800
cgtgctgaag	ggcgtgaagc	tgactacac	ctgataactc	gagaattcac	gcgtggtacc	4860
tctagagtcg	acccggggcg	ccgcttcgag	cagacatgat	aagatacatt	gatgagtttg	4920
gacaaaccac	aactagaatg	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	4980
ttgctttatt	tgtaaccatt	ataagctgca	ataaacaagt	taacaacaac	aattgcattc	5040

S226CAS108.ST25

attttatggt tcagggttcag ggggagatgt gggagggtttt ttaaagcaag taaaacctct 5100
 acaaagtgtg taaaatcgat aaggatccgg gctggcgtaa tagcgaagag gcccgcacccg 5160
 atcgcccttc ccaacagttg cgcagcctga atggcgaatg gacgcgccct gtagcggcgc 5220
 attaagcgcg gcgggtgtgg tggttacgcg cagcgtgacc gctacacttg ccagcgccct 5280
 agcgcccgt cctttcgctt tcttcccttc ctttctcgcc acgttcgccg gctttccccg 5340
 tcaagctcta aatcgggggc tccctttagg gttccgattt agagctttac ggcacctcga 5400
 ccgcaaaaaa cttgatttgg gtgatgggtt acgtagtggg ccatcgccct gatagacggt 5460
 ttttcgccct ttgacgttgg agtccacgtt ctttaatatg ggactcttgt tccaaactgg 5520
 aacaacactc aaccctatct cggctctattc ttttgattta taagggtatt tgccgatttc 5580
 ggcctatttg ttaaaaaatg agctgattta acaaataatt aacgcgaatt ttaacaaaat 5640
 attaacgttt acaatttcgc ctgatgcggt attttctcct tacgcatctg tgcggtattt 5700
 cacaccgat atggtgcact ctcagtacaa tctgctctga tgccgcatag ttaagccagc 5760
 cccgacaccc gccaacaccc gctgacgcgc cctgacgggc ttgtctgctc ccggcatccg 5820
 ctacagaca agctgtgacc gtctccggga gctgcatgtg tcagagggtt tcaccgtcat 5880
 caccgaaacg cgcgagacga aagggcctcg tgatacgccct atttttatag gttaatgtca 5940
 tgataataat ggtttcttag acgtcaggtg gcacttttcg gggaaatgtg cgcggaaccc 6000
 ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga caataaccct 6060
 gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat ttccgtgtcg 6120
 cccttattcc cttttttgcg gcattttgcc ttcctgtttt tgctcaccca gaaacgctgg 6180
 tgaaagtaaa agatgctgaa gatcagttgg gtgcaċgagt gggttacatc gaactggatc 6240
 tcaacagcgg taagatcctt gagagttttc gccccgaaga acgttttcca atgatgagca 6300
 cttttaaagt tctgctatgt ggcgcggtat tatcccgtat tgacgccggg caagagcaac 6360
 tcggtcgccg catacactat tctcagaatg acttggttga gtactcacca gtcacagaaa 6420
 agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata accatgagt 6480
 ataacactgc ggccaactta cttctgacaa cgatcgagg accgaaggag ctaaccgctt 6540
 ttttgacaaa catgggggat catgtaactc gccttgatcg ttgggaaccg gagctgaatg 6600
 aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca acaacgttgc 6660
 gcaaactatt aactggcgaa ctacttactc tagcttcccg gcaacaatta atagactgga 6720
 tggaggcgga taaagttgca ggaccacttc tgcgctcggc ccttccggct ggctggttta 6780
 ttgctgataa atctggagcc ggtgagcgtg ggtctcgcgg tatcattgca gcactggggc 6840
 cagatggtaa gccctcccgt atcgtagtta tctacacgac ggggagtcag gcaactatgg 6900
 atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat tggtaactgt 6960
 cagaccaagt ttactcatat atactttaga ttgatttaaa acttcatttt taatttaaaa 7020
 ggatctaggt gaagatcctt ttgataatc tcatgaccaa aatcccttaa cgtgagtttt 7080

S226CAS108.ST25

```

cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga gatccttttt 7140
ttctgcgcgt aatctgctgc ttgcaaacaa aaaaaccacc gctaccagcg gtggtttggt 7200
tgccggatca agagctacca actctttttt cgaaggtaac tggcttcagc agagcgcaga 7260
taccaaatac tgtccttcta gtgtagccgt agttaggccca ccacttcaag aactctgtag 7320
caccgcctac atacctcgct ctgctaatac tgttaccagt ggctgctgcc agtggcgata 7380
agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg cagcggtcgg 7440
gctgaacggg gggttcgtgc acacagccca gcttggagcg aacgacctac accgaactga 7500
gatactaca gcgtgagcta tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca 7560
ggtatccggt aagcggcagg gtcggaacag gagagcgcac gagggagctt ccagggggaa 7620
acgcctggtg tctttatagt cctgtcgggt ttcgccacct ctgacttgag cgtcgatttt 7680
tgtgatgctc gtcagggggg cggagcctat ggaaaaacgc cagcaacgcg gcctttttac 7740
ggttcctggc cttttgctgg ctttttgctc acatggctcg acagatct 7788

```

<210> 141

<211> 23

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SNE-S1

<400> 141

ggttgggatt atccaaaatg tga

23

<210> 142

<211> 24

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SNE-AS1

<400> 142

gcatcatcag aaagaatcat catg

24

<210> 143

<211> 21

<212> DNA

S226CAS108.ST25

<213> Artificial sequence

<220>

<223> amorce SAR1-S

<400> 143

cctctcttgt tcttgctcgc a

21

<210> 144

<211> 21

<212> DNA

<213> Artificial sequence

<220>

<223> amorce SAR1-AS

<400> 144

tatagtgagc cgccacacat g

21

<210> 145

<211> 45

<212> DNA

<213> Artificial sequence

<220>

<223> amorce PCR

<400> 145

ataggatcca ccatgtttat tttcttatta tttcttactc tcact

45

<210> 146

<211> 37

<212> DNA

<213> Artificial sequence

<220>

<223> amorce PCR

<400> 146

atactcgagt tatgtgtaat gtaatttgac acccttg

37

<210> 147

S226CAS108.ST25

<211> 45

<212> DNA

<213> Artificial sequence

<220>

<223> amorce PCR

<400> 147

ataggatcca ccatgtttat tttcttatta tttcttactc tcact

45

<210> 148

<211> 36

<212> DNA

<213> Artificial sequence

<220>

<223> amorce PCR

<400> 148

acctccggat ttaatatatt gtcataattt tcccaa

36

<210> 149

<211> 13

<212> PRT

<213> Artificial sequence

<220>

<223> extrémité N-terminale de la protéine S du SRAS-CoV (acides amines
1 à 13)

<400> 149

Met Phe Ile Phe Leu Leu Phe Leu Thr Leu Thr Ser Gly
1 5 10

<210> 150

<211> 10

<212> PRT

<213> Artificial sequence

<220>

<223> oligopeptide

S226CAS108.ST25

<400> 150

Ser Gly Asp Tyr Lys Asp Asp Asp Asp Lys
1 5 10

<210> 151

<211> 34

<212> DNA

<213> Artificial sequence

<220>

<223> amorce PCR

<400> 151

actagctagc ggatccacca tggtcatctt cctg

34

<210> 152

<211> 33

<212> DNA

<213> Artificial sequence

<220>

<223> amorce PCR

<400> 152

agtatccgga cttgatgtac tgctcgtagt tgc

33

<210> 153

<211> 59

<212> DNA

<213> Artificial sequence

<220>

<223> oligonucléotide

<400> 153

tatgagcttt tttttttttt ttttttggc atataaatag actcggcgcg ccatctgca

59

<210> 154

<211> 53

<212> DNA

<213> Artificial sequence

S226CAS108.ST25

<220>

<223> oligonucléotide

<400> 154

gatggcgcgc cgagtctatt tatatgccaa aaaaaaaaaa aaaaaaaagc tca 53

<210> 155

<211> 45

<212> DNA

<213> Artificial sequence

<220>

<223> amorce PCR

<400> 155

atacgtacga ccatgtttat tttcttatta tttcttactc tcact 45

<210> 156

<211> 40

<212> DNA

<213> Artificial sequence

<220>

<223> amorce PCR

<400> 156

atagcgcgct cattatgtgt aatgtaattt gacacccttg 40

<210> 157

<211> 20

<212> DNA

<213> Artificial sequence

<220>

<223> amorce PCR

<400> 157

ccatttcaac aatttggccg 20

<210> 158

<211> 45

S226CAS108.ST25

<212> DNA

<213> Artificial sequence

<220>

<223> amorce PCR

<400> 158

ataggatccg cgcgctcatt atttatcgtc gtcattcttta taatc

45

0-1	Formulaire PCT/RO/134 (SAFE) Indications relatives à des micro-organismes ou autre matériel biologique déposés (règle 13bis du PCT)	
0-1-1	Préparée avec	PCT-SAFE [EASY mode] Version 3.50 (Build 0002.162)
0-2	Demande internationale n°	
0-3	Référence du dossier du déposant ou du mandataire	VMACP226-108
1	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
1-1	page	11
1-2	ligne	31
1-3	Identification du dépôt	
1-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
1-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
1-3-3	Date du dépôt	20 Juin 2003 (20.06.2003)
1-3-4	Numéro d'ordre	CNCM I-3059
1-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
1-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
2	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
2-1	page	12
2-2	ligne	6
2-3	Identification du dépôt	
2-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
2-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
2-3-3	Date du dépôt	12 Mai 2003 (12.05.2003)
2-3-4	Numéro d'ordre	CNCM I-3020
2-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
2-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

3	Les Indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
3-1	page	12
3-2	ligne	13
3-3	Identification du dépôt	
3-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
3-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
3-3-3	Date du dépôt	12 Mai 2003 (12.05.2003)
3-3-4	Numéro d'ordre	CNCM I-3019
3-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
3-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
4	Les Indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
4-1	page	12
4-2	ligne	20
4-3	Identification du dépôt	
4-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
4-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
4-3-3	Date du dépôt	13 Novembre 2003 (13.11.2003)
4-3-4	Numéro d'ordre	CNCM I-3126
4-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
4-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

5	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
5-1	page	12
5-2	ligne	28
5-3	Identification du dépôt	
5-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
5-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
5-3-3	Date du dépôt	28 Mai 2003 (28.05.2003)
5-3-4	Numéro d'ordre	CNCM I-3046
5-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
5-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
6	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
6-1	page	13
6-2	ligne	4
6-3	Identification du dépôt	
6-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
6-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
6-3-3	Date du dépôt	28 Mai 2003 (28.05.2003)
6-3-4	Numéro d'ordre	CNCM I-3047
6-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
6-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

7	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
7-1	page	13
7-2	ligne	11
7-3	Identification du dépôt	
7-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
7-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
7-3-3	Date du dépôt	13 Novembre 2003 (13.11.2003)
7-3-4	Numéro d'ordre	CNCM I-3125
7-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
7-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
8	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
8-1	page	13
8-2	ligne	19
8-3	Identification du dépôt	
8-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
8-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
8-3-3	Date du dépôt	05 Juin 2003 (05.06.2003)
8-3-4	Numéro d'ordre	CNCM I-3048
8-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
8-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

9	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
9-1	page	13
9-2	ligne	30
9-3	Identification du dépôt	
9-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
9-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
9-3-3	Date du dépôt	07 Novembre 2003 (07.11.2003)
9-3-4	Numéro d'ordre	CNCM I-3124
9-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
9-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
10	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
10-1	page	14
10-2	ligne	5
10-3	Identification du dépôt	
10-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
10-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
10-3-3	Date du dépôt	07 Novembre 2003 (07.11.2003)
10-3-4	Numéro d'ordre	CNCM I-3123
10-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
10-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

11	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
11-1	page	15
11-2	ligne	5
11-3	Identification du dépôt	
11-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
11-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
11-3-3	Date du dépôt	23 Octobre 2003 (23.10.2003)
11-3-4	Numéro d'ordre	CNCM I-3117
11-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
11-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
12	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
12-1	page	15
12-2	ligne	9
12-3	Identification du dépôt	
12-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
12-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
12-3-3	Date du dépôt	23 Octobre 2003 (23.10.2003)
12-3-4	Numéro d'ordre	CNCM I-3118
12-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
12-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

13	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
13-1	page	15
13-2	ligne	18
13-3	Identification du dépôt	
13-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
13-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
13-3-3	Date du dépôt	22 Novembre 2004 (22.11.2004)
13-3-4	Numéro d'ordre	CNCM I-3323
13-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
13-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
14	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
14-1	page	15
14-2	ligne	19
14-3	Identification du dépôt	
14-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
14-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
14-3-3	Date du dépôt	22 Novembre 2004 (22.11.2004)
14-3-4	Numéro d'ordre	CNCM I-3324
14-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
14-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

15	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
15-1	page	17
15-2	ligne	27
15-3	Identification du dépôt	
15-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
15-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
15-3-3	Date du dépôt	22 Novembre 2004 (22.11.2004)
15-3-4	Numéro d'ordre	CNCM I-3325
15-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
15-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
16	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
16-1	page	15
16-2	ligne	20
16-3	Identification du dépôt	
16-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
16-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
16-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
16-3-4	Numéro d'ordre	CNCM I-3326
16-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
16-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

17	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
17-1	page	15
17-2	ligne	21
17-3	Identification du dépôt	
17-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
17-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
17-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
17-3-4	Numéro d'ordre	CNCM I-3327
17-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
17-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
18	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
18-1	page	19
18-2	ligne	28
18-3	Identification du dépôt	
18-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
18-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
18-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
18-3-4	Numéro d'ordre	CNCM I-3328
18-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
18-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

19	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
19-1	page	19
19-2	ligne	30
19-3	Identification du dépôt	
19-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
19-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
19-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
19-3-4	Numéro d'ordre	CNCM I-3329
19-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
19-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
20	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
20-1	page	19
20-2	ligne	32
20-3	Identification du dépôt	
20-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
20-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
20-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
20-3-4	Numéro d'ordre	CNCM I-3330
20-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
20-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

21	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
21-1	page	20
21-2	ligne	2
21-3	Identification du dépôt	
21-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
21-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
21-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
21-3-4	Numéro d'ordre	CNCM I-3331
21-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
21-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
22	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
22-1	page	15
22-2	ligne	22
22-3	Identification du dépôt	
22-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
22-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
22-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
22-3-4	Numéro d'ordre	CNCM I-3332
22-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
22-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

23	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
23-1	page	15
23-2	ligne	23
23-3	Identification du dépôt	
23-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
23-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
23-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
23-3-4	Numéro d'ordre	CNCM I-3333
23-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
23-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
24	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
24-1	page	15
24-2	ligne	24
24-3	Identification du dépôt	
24-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
24-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
24-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
24-3-4	Numéro d'ordre	CNCM I-3334
24-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
24-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

25	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
25-1	page	15
25-2	ligne	25
25-3	Identification du dépôt	
25-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
25-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
25-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
25-3-4	Numéro d'ordre	CNCM I-3335
25-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
25-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
26	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
26-1	page	15
26-2	ligne	26
26-3	Identification du dépôt	
26-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
26-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
26-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
26-3-4	Numéro d'ordre	CNCM I-3336
26-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
26-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

27	Les Indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
27-1	page	15
27-2	ligne	27
27-3	Identification du dépôt	
27-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
27-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
27-3-3	Date du dépôt	01 Decembre 2004 (01.12.2004)
27-3-4	Numéro d'ordre	CNCM I-3337
27-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
27-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
28	Les Indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
28-1	page	15
28-2	ligne	28
28-3	Identification du dépôt	
28-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
28-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
28-3-3	Date du dépôt	02 Decembre 2004 (02.12.2004)
28-3-4	Numéro d'ordre	CNCM I-3338
28-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
28-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

29	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
29-1	page	15
29-2	ligne	29
29-3	Identification du dépôt	
29-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
29-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
29-3-3	Date du dépôt	02 Decembre 2004 (02.12.2004)
29-3-4	Numéro d'ordre	CNCM I-3339
29-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
29-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations
30	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
30-1	page	15
30-2	ligne	30
30-3	Identification du dépôt	
30-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
30-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
30-3-3	Date du dépôt	02 Decembre 2004 (02.12.2004)
30-3-4	Numéro d'ordre	CNCM I-3340
30-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
30-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

31	Les indications ci-dessous ont trait au micro-organisme ou autre matériel biologique visé dans la description :	
31-1	page	15
31-2	ligne	31
31-3	Identification du dépôt	
31-3-1	Nom de l'institution de dépôt	CNCM Collection nationale de cultures de micro-organismes
31-3-2	Adresse de l'institution de dépôt	FR- Institut Pasteur, 28, rue du Dr Roux, F-75724 Paris Cedex 15
31-3-3	Date du dépôt	02 Decembre 2004 (02.12.2004)
31-3-4	Numéro d'ordre	CNCM I-3341
31-4	Indications supplémentaires	En ce qui concerne les désignations dans lesquelles un Brevet est demandé, le micro-organisme déposé ne sera accessible que par la remise d'un échantillon à un expert désigné par le requérant.
31-5	Etats désignés pour lesquels les indications sont données	de toutes les désignations

RÉSERVÉ À L'OFFICE RÉCEPTEUR

0-4	Cette feuille a été reçue en même temps que la demande internationale : (oui ou non)	OUI
0-4-1	Fonctionnaire autorisé	Lionel BUFFET

RÉSERVÉ AU BUREAU INTERNATIONAL

0-5	Cette feuille est parvenue au Bureau international le :	
0-5-1	Fonctionnaire autorisé	

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR2004/003106

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7	C12N7/00	C12N15/50	C12N15/63	C07K14/165	C07K16/10
	C12Q1/68	C12Q1/70	G01N33/53	G01N33/569	C07H21/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C07K C12Q G01N C07H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, EMBASE, BIOSIS, Sequence Search

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL 22 April 2003 (2003-04-22), XP002294758 Database accession no. AY278489 the whole document	5-29
X	DATABASE EMBL 10 June 2003 (2003-06-10), XP002294760 Database accession no. AY290752 the whole document	5,7-28
X	DATABASE UNIPROT 10 October 2003 (2003-10-10), XP002294761 Database accession no. P59595 the whole document	5,7-28
	----- -/--	



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

* Special categories of cited documents:

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *G* document member of the same patent family

Date of the actual completion of the international search

2 June 2005

Date of mailing of the international search report

01/07/2005

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Chavanne, F

INTERNATIONAL SEARCH REPORT

International Application No

PCT/FR2004/003106

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MARRA M A ET AL: "The genome sequence of the SARS-associated coronavirus" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 300, no. 5624, 30 May 2003 (2003-05-30), pages 1399-1404, XP002269483 ISSN: 0036-8075 the whole document	5-29
X	CHE X-Y ET AL: "RAPID AND EFFICIENT PREPARATION OF MONOCLONAL ANTIBODIES AGAINST SARS-ASSOCIATED CORONAVIRUS NUCLEOCAPSID PROTEIN BY IMMUNIZING MICE" DI YI JUNYI DAXUE XUEBAO - ACADEMIC JOURNAL OF FIRST MEDICAL COLLEGE OF PLA, GAI KAN BIANJISHI, GUANGZHOU, CN, vol. 23, no. 7, July 2003 (2003-07), pages 640-642, XP008028243 ISSN: 1000-2588 the whole document	5-29
X	WANG J ET AL: "ASSESSMENT OF IMMUNOREACTIVE SYNTHETIC PEPTIDES FROM THE STRUCTURAL PROTEINS OF SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS" CLINICAL CHEMISTRY, AMERICAN ASSOCIATION FOR CLINICAL CHEMISTRY. WINSTON, US, vol. 49, no. 12, 13 November 2003 (2003-11-13), pages 1989-1996, XP001182489 ISSN: 0009-9147 the whole document	5-29
X,P	SHI Y ET AL: "DIAGNOSIS OF SEVERE ACUTE RESPIRATORY SYNDROME (SARS) BY DETECTION OF SARS CORONAVIRUS NUCLEOCAPSID ANTIBODIES IN AN ANTIGEN-CAPTURING ENZYME-LINKED IMMUNOSORBENT ASSAY" JOURNAL OF CLINICAL MICROBIOLOGY, WASHINGTON, DC, US, vol. 41, no. 12, December 2003 (2003-12), pages 5781-5782, XP008028263 ISSN: 0095-1137 the whole document	7-24,27, 28
X,P	LIU G ET AL: "The C-Terminal Portion of the Nucleocapsid Protein Demonstrates SARS-CoV Antigenicity" GENOMICS, PROTEOMICS AND BIOINFORMATICS, vol. 1, no. 3, 2003, pages 193-197, XP001183377 the whole document	7-10,21, 22,28

-/--

INTERNATIONAL SEARCH REPORT

International Application No
PCT/FR2004/003106

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>POON LEO L M ET AL: "Rapid diagnosis of a coronavirus associated with severe acute respiratory syndrome (SARS)" CLINICAL CHEMISTRY, AMERICAN ASSOCIATION FOR CLINICAL CHEMISTRY. WINSTON, US, vol. 49, no. 6 Pt 1, June 2003 (2003-06), pages 953-955, XP002288942 ISSN: 0009-9147</p> <p>-----</p>	

RAPPORT DE RECHERCHE INTERNATIONALE

Demande internationale No
PCT/FR2004/003106

A. CLASSEMENT DE L'OBJET DE LA DEMANDE CIB 7 C12N7/00 C12N15/50 C12N15/63 C07K14/165 C07K16/10 C12Q1/68 C12Q1/70 G01N33/53 G01N33/569 C07H21/00			
Selon la classification internationale des brevets (CIB) ou à la fois selon la classification nationale et la CIB			
B. DOMAINES SUR LESQUELS LA RECHERCHE A PORTE Documentation minimale consultée (système de classification suivi des symboles de classement) CIB 7 C12N C07K C12Q G01N C07H			
Documentation consultée autre que la documentation minimale dans la mesure où ces documents relèvent des domaines sur lesquels a porté la recherche			
Base de données électronique consultée au cours de la recherche internationale (nom de la base de données, et si réalisable, termes de recherche utilisés) EPO-Internal, WPI Data, PAJ, EMBASE, BIOSIS, Sequence Search			
C. DOCUMENTS CONSIDERES COMME PERTINENTS			
Catégorie °	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées	
X	DATABASE EMBL 22 avril 2003 (2003-04-22), XP002294758 Database accession no. AY278489 le document en entier	5-29	
X	DATABASE EMBL 10 juin 2003 (2003-06-10), XP002294760 Database accession no. AY290752 le document en entier	5,7-28	
X	DATABASE UNIPROT 10 octobre 2003 (2003-10-10), XP002294761 Database accession no. P59595 le document en entier	5,7-28	
-/-			
<div style="display: flex; justify-content: space-between;"> <input checked="" type="checkbox"/> Voir la suite du cadre C pour la fin de la liste des documents <input type="checkbox"/> Les documents de familles de brevets sont indiqués en annexe </div>			
<div style="display: flex; justify-content: space-between;"> <div style="width: 48%;"> <p>° Catégories spéciales de documents cités:</p> <p>*A* document définissant l'état général de la technique, non considéré comme particulièrement pertinent</p> <p>*E* document antérieur, mais publié à la date de dépôt international ou après cette date</p> <p>*L* document pouvant jeter un doute sur une revendication de priorité ou cité pour déterminer la date de publication d'une autre citation ou pour une raison spéciale (telle qu'indiquée)</p> <p>*O* document se référant à une divulgation orale, à un usage, à une exposition ou tous autres moyens</p> <p>*P* document publié avant la date de dépôt international, mais postérieurement à la date de priorité revendiquée</p> </div> <div style="width: 48%;"> <p>*T* document ultérieur publié après la date de dépôt international ou la date de priorité et n'appartenant pas à l'état de la technique pertinent, mais cité pour comprendre le principe ou la théorie constituant la base de l'invention</p> <p>*X* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme nouvelle ou comme impliquant une activité inventive par rapport au document considéré isolément</p> <p>*Y* document particulièrement pertinent; l'invention revendiquée ne peut être considérée comme impliquant une activité inventive lorsque le document est associé à un ou plusieurs autres documents de même nature, cette combinaison étant évidente pour une personne du métier</p> <p>*&* document qui fait partie de la même famille de brevets</p> </div> </div>			
Date à laquelle la recherche internationale a été effectivement achevée <div style="text-align: center; font-weight: bold;">2 juin 2005</div>		Date d'expédition du présent rapport de recherche internationale <div style="text-align: center; font-weight: bold;">01/07/2005</div>	
Nom et adresse postale de l'administration chargée de la recherche internationale Office Européen des Brevets, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Fonctionnaire autorisé <div style="text-align: center; font-weight: bold;">Chavanne, F</div>	

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
X	MARRA M A ET AL: "The genome sequence of the SARS-associated coronavirus" SCIENCE, AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE,, US, vol. 300, no. 5624, 30 mai 2003 (2003-05-30), pages 1399-1404, XP002269483 ISSN: 0036-8075 le document en entier	5-29
X	CHE X-Y ET AL: "RAPID AND EFFICIENT PREPARATION OF MONOCLONAL ANTIBODIES AGAINST SARS-ASSOCIATED CORONAVIRUS NUCLEOCAPSID PROTEIN BY IMMUNIZING MICE" DI YI JUNYI DAXUE XUEBAO - ACADEMIC JOURNAL OF FIRST MEDICAL COLLEGE OF PLA, GAI KAN BIANJISHI, GUANGZHOU, CN, vol. 23, no. 7, juillet 2003 (2003-07), pages 640-642, XP008028243 ISSN: 1000-2588 le document en entier	5-29
X	WANG J ET AL: "ASSESSMENT OF IMMUNOREACTIVE SYNTHETIC PEPTIDES FROM THE STRUCTURAL PROTEINS OF SEVERE ACUTE RESPIRATORY SYNDROME CORONAVIRUS" CLINICAL CHEMISTRY, AMERICAN ASSOCIATION FOR CLINICAL CHEMISTRY. WINSTON, US, vol. 49, no. 12, 13 novembre 2003 (2003-11-13), pages 1989-1996, XP001182489 ISSN: 0009-9147 le document en entier	5-29
X,P	SHI Y ET AL: "DIAGNOSIS OF SEVERE ACUTE RESPIRATORY SYNDROME (SARS) BY DETECTION OF SARS CORONAVIRUS NUCLEOCAPSID ANTIBODIES IN AN ANTIGEN-CAPTURING ENZYME-LINKED IMMUNOSORBENT ASSAY" JOURNAL OF CLINICAL MICROBIOLOGY, WASHINGTON, DC, US, vol. 41, no. 12, décembre 2003 (2003-12), pages 5781-5782, XP008028263 ISSN: 0095-1137 le document en entier	7-24,27, 28
X,P	LIU G ET AL: "The C-Terminal Portion of the Nucleocapsid Protein Demonstrates SARS-CoV Antigenicity" GENOMICS, PROTEOMICS AND BIOINFORMATICS, vol. 1, no. 3, 2003, pages 193-197, XP001183377 le document en entier	7-10,21, 22,28

-/--

C.(suite) DOCUMENTS CONSIDERES COMME PERTINENTS		
Catégorie	Identification des documents cités, avec, le cas échéant, l'indication des passages pertinents	no. des revendications visées
A	<p>POON LEO L M ET AL: "Rapid diagnosis of a coronavirus associated with severe acute respiratory syndrome (SARS)" CLINICAL CHEMISTRY, AMERICAN ASSOCIATION FOR CLINICAL CHEMISTRY. WINSTON, US, vol. 49, no. 6 Pt 1, juin 2003 (2003-06), pages 953-955, XP002288942 ISSN: 0009-9147</p> <p>-----</p>	